

Modellbasierte Regelung von Nanopositionier- und Nanomessmaschinen

**Dissertation
zur Erlangung des akademischen Grades
Doktoringenieur (Dr.-Ing.)**

**vorgelegt der Fakultät für Informatik und Automatisierung
der Technischen Universität Ilmenau**

**von Dipl.-Ing. Arvid Amthor
geboren am 30. Juli 1979 in Meiningen**

Gutachter:

Herr Univ.-Prof. Dr.-Ing. habil. Christoph Ament, Technische Universität Ilmenau

Herr Prof. Dr.-Ing. Steven Lambeck (JP), Technische Universität Ilmenau

Herr Univ.-Prof. Dr.-Ing. habil. Oliver Sawodny, Universität Stuttgart

vorgelegt am: 02.02.2010

verteidigt am: 24.09.2010

Vorwort

Die vorliegende Dissertation wurde während meiner vierjährigen Tätigkeit als wissenschaftlicher Mitarbeiter am Fachgebiet Systemanalyse der Technischen Universität Ilmenau angefertigt. Sie entstand im Rahmen des Sonderforschungsbereiches „Nanopositionier- und Nanomessmaschinen“. Dass die in dieser Zeit durchgeführten Forschungsarbeiten so fruchtbar waren, um schließlich in der folgenden Arbeit zusammengefasst zu werden, ist der tatkräftigen Unterstützung von Seiten Vorgesetzter, Kollegen, Studenten und insbesondere dem Beistand durch die eigene Familie zu verdanken.

Zu Beginn möchte ich mich bei meinem Doktorvater Prof. Dr.-Ing. habil. Christoph Ament für die hervorragende wissenschaftliche Betreuung der Arbeit bedanken. Auch möchte ich Herrn Prof. Dr.-Ing. habil. Gerd Jäger meinen besonderen Dank aussprechen, denn er gab mir die Möglichkeit, am Sonderforschungsbereich 622 mitzuarbeiten.

Ebenso möchte ich meinen Kollegen Prof. Dr.-Ing. Steven Lambeck (JP), Dipl.-Ing. René Güttler, Dipl.-Ing. Stephan Zschäck, Dipl.-Ing. Tran-Trung Nguyen, Dipl.-Ing. Eberhard Gebel, Volker Pranner und Mandy Assinner-Hoffmann für die sowohl fachliche als auch persönliche Unterstützung danken. Sie haben mich durch die vielen fruchtbaren und anregenden Diskussionen stets motiviert und somit wesentlich zum Entstehen dieser Arbeit beigetragen.

Des Weiteren möchte ich mich bei den Mitarbeitern des Sonderforschungsbereiches 622 bedanken. Durch die stets konstruktive Zusammenarbeit und das ausgesprochen angenehme Arbeitsklima trugen sie in einem erheblichen Maße zum Gelingen der vorliegenden Arbeit bei. Mein besonderer Dank gilt hierbei Dr.-Ing. Tino Hausotte für die wertvollen Diskussionen und Hilfestellungen während meiner Tätigkeit.

Weiterhin bedanke ich mich bei den von mir betreuten Diplomanden, Studienarbeitern und studentischen Hilfskräften, welche mich durch ihre unermüdliche Arbeit bei der Anfertigung dieser Arbeit unterstützt haben.

Für die gewinnbringenden fachlichen Diskussionen bei der Durchsicht des Manuskriptes danke ich Dr.-Ing. Mike Eichhorn sowie allen weiteren Korrekturlesern.

Durch ihre Unterstützung und ihr Verständnis trug meine Lebensgefährtin Maria entscheidend zum Gelingen dieser Arbeit bei. Hierfür danke ich dir von Herzen. Meine besondere Anerkennung gilt auch meiner Mutter für die immerwährende fürsorgliche Unterstützung, ihren Beistand und ihre aufmunternden Worte.

Für meine Tochter

Nele

...

Inhaltsverzeichnis

1	Einleitung	1
1.1	Die Nanomessmaschine 1	2
1.2	Zielstellung der Arbeit	6
1.3	Aufbau der Arbeit	7
2	Experimenteller Aufbau.....	9
2.1	Mechanischer Aufbau	11
2.2	Sensorsystem.....	12
2.3	Antriebssystem.....	14
2.4	Positionsinitialisierung.....	16
2.5	Echtzeithardware.....	18
3	Modellbildung	21
3.1	Dynamisches Verhalten des mechanischen Aufbaus.....	21
3.2	Dynamisches Verhalten des Sensorsystems	22
3.3	Dynamisches Verhalten des Aktorsystems.....	22
3.4	Modellierung der Reibkraft.....	25
3.4.1	Reibphänomene	26
3.4.2	Stand der Technik auf dem Gebiet der Reibmodellierung	32
3.4.2.1	Physikalisch motivierte Reibmodelle	32
3.4.2.2	Datenbasierte Reibmodelle.....	41
3.4.3	Parameteridentifikation	43
3.4.3.1	Identifikation des LuGre- und des elasto-plastischen Reibmodells	44
3.4.3.2	Identifikation des GMS-Modells.....	48
3.4.3.3	Identifikation des DNLRX-Modells.....	51
3.4.3.4	Identifikation des künstlichen Neuronalen Netzes	52
3.4.4	Leistungsfähigkeit der identifizierten Reibmodelle	56
3.4.4.1	Ergebnisse der Validierung	57
3.4.4.2	Leistungsbewertung.....	60
4	Dynamisches Regelungskonzept.....	62
4.1	Bahnplanung	63
4.1.1	Begriffsbestimmung	63
4.1.2	Grundlagen der analytischen Bahnplanung.....	64
4.1.2.1	Zusammenhang zwischen Bahn- und Trajektorienkinematik	65
4.1.2.2	Bestimmung der Kurvenkinematik für die Strecke.....	67

4.1.2.3	Bestimmung der Kurvenkinematik für den Kreisbogen.....	69
4.1.3	Bestimmung der Restriktionen auf Bahnebene	73
4.1.3.1	Bestimmung der Bahnrestriktionen für die Strecke	73
4.1.3.2	Bestimmung der Bahnrestriktionen für den Kreisbogen.....	74
4.1.4	Kinematikplanung auf Bahnebene	81
4.1.4.1	Stand der Technik auf dem Gebiet der analytischen Bahnplanung.....	81
4.1.4.2	Bahnplanung vierter Ordnung	83
4.1.5	Kontinuierliche Bahnplanung mittels Verschleifelementen.....	88
4.1.5.1	Bestimmung der Kurvenkinematik für die erweiterte Blosskurve	90
4.1.5.2	Bestimmung der Bahnrestriktionen für die erweiterte Blosskurve	91
4.1.5.3	Bestimmung der Übergangsgeschwindigkeiten	93
4.2	Unterlagerte Regelung	97
4.2.1	PID-Regler.....	98
4.2.2	PI-Zustandsregelung.....	98
4.2.2.1	Entwurf mittels Polvorgabe	98
4.2.2.2	Entwurf einer Vorsteuerung mit Führungsgrößenaufschaltung	101
4.2.3	Zustandsrekonstruktion	103
4.3	Adaptive Vorsteuerung	105
4.3.1	Adaptives DNLRX Modell	106
4.3.2	Adaptives neuronales Netz	111
4.4	Störbeobachtung	113
4.4.1	Prinzipielle Funktionsweise eines Kalman Filters	114
4.4.2	Störbeobachtung in NPM-Systemen	115
5	Leistungsfähigkeit des dynamischen Regelungskonzeptes	117
5.1	Untersuchte Trajektorien	117
5.2	Einfluss der Regelungskomponenten auf die erzielbare Regelgüte.....	118
5.3	Bewertung der Ergebnisse	120
6	Regelung von X- und Y- Achse.....	123
6.1	Dezentrales Regelungskonzept	123
6.2	Experimentelle Untersuchungen.....	124
6.2.1	Kreise in der xy-Ebene	124
6.2.2	Komplexe Figuren in der xy-Ebene	127
7	Zusammenfassung und Ausblick	131
Anhang	134	
Anhang A.....	134	
Anhang B.....	139	
Anhang C.....	150	
Anhang D.....	157	
Literaturverzeichnis.....	162	

Variablen und Abkürzungen

Variablen

$(\cdot)'$	Ableitung nach dem Bahnort
$\dot{(\cdot)}$	Ableitung nach der Zeit
$\tilde{(\cdot)}$	Transformierte Größe
$\hat{(\cdot)}$	Abschätzung einer Größe
$(\cdot)_{opt}$	Optimum einer Größe
$\mathbf{0}$	Nullvektor
\mathbf{A}	Dynamikmatrix
$a, a(t)$	(Bahn-) Beschleunigungsfunktion
$\vec{a}, \vec{a}(t)$	(Kurven-) Beschleunigungsfunktion, $\vec{a}(t) = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix}$
a_{max}	(Bahn-) Beschleunigungsrestriktion
\vec{a}_{max}	(Kurven-) Beschleunigungsrestriktion, $\vec{a}_{max}(t) = \begin{pmatrix} a_{max,x} \\ a_{max,y} \\ a_{max,z} \end{pmatrix}$
$a_{max,x}, a_{max,y}, a_{max,z}$	Beschleunigungsrestriktion der x-, y-, z-Achse
$\mathbf{A_R}$	Dynamikmatrix des geregelten Systems
a_x, a_y, a_z	Beschleunigungsfunktion der x-, y-, z-Achse
b	Offset des DNLRX-Reibmodells
$\vec{b}, \vec{b}(s)$	(normierter) Binormalenvektor
B	Magnetische Flussdichte
\mathbf{B}	Steuermatrix

c	Federkonstante
\mathbf{C}	Ausgangsmatrix
d	Dämpfungskonstante
$\mathbf{D}(k)$	Diagonalmatrix bei der U-D-Faktorisierung
$d, d(t)$	(Bahn-) Ruckanstiegsfunktion
$\vec{d}, \vec{d}(t)$	(Kurven-) Ruckanstiegsfunktion, $\vec{d}(t) = \begin{pmatrix} d_x \\ d_y \\ d_z \end{pmatrix}$
d_{max}	(Bahn-) Ruckanstiegsrestriktion
\vec{d}_{max}	(Kurven-) Ruckanstiegsrestriktion, $\vec{d}_{max}(t) = \begin{pmatrix} d_{max,x} \\ d_{max,y} \\ d_{max,z} \end{pmatrix}$
$d_{max,x}, d_{max,y}, d_{max,z}$	Ruckanstiegsrestriktion der x-, y-, z-Achse
D_V	Intervallgrenze
d_x, d_y, d_z	Ruckanstiegsfunktion der x-, y-, z-Achse
\mathbf{E}	Vektor der Ausgangsfehler des KNN
$e(t)$	Regelfehler
$\mathbf{e}_M(k)$	Modellfehler
F	Angreifende Kraft
$F_A(t)$	Antriebskraft
F_C	Coulombkraft
$F_E(t)$	Elektrodynamische Kraft
f_g	Knickfrequenz
$F_i(t)$	Kraftbeitrag eines „Maxwell-Slip“-Elementes
$F_{LM}(t)$	Antriebskraft eines Linearmotors
F_R	Stationäre Reibkraft
$F_R(t)$	Dynamische Reibkraft
F_S	Haftreibungskraft
$F_S(t)$	Kraft, die von Störungen erzeugt wird
$F_{ss}(\dot{x}(t))$	Funktion der Reibkraft im stationären Fall (Stribeckkurve)
\mathbf{I}	Einheitsmatrix

$i(t)$	Strom
$i_A(t)$	Antriebsstrom
J	Gütefunktional
\mathbf{J}	Jakobimatrix
J	Gütefunktion
$j, j(t)$	(Bahn-) Ruckfunktion
$\vec{j}, \vec{j}(t)$	(Kurven-) Ruckfunktion, $\vec{j}(t) = \begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix}$
j_{max}	(Bahn-) Ruckrestriktion
\vec{j}_{max}	(Kurven-) Ruckrestriktion, $\vec{j}_{max}(t) = \begin{pmatrix} j_{max, x} \\ j_{max, y} \\ j_{max, z} \end{pmatrix}$
$j_{max, x}, j_{max, y}, j_{max, z}$	Ruckrestriktion der x-, y-, z-Achse
j_x, j_y, j_z	Ruckfunktion der x-, y-, z-Achse
\mathbf{K}	Verstärkungsmatrix Zustandsregler
k_A	Akkumulierte Kraftkonstante des Antriebs einer Achse
K_D	Verstärkung Differentialanteil
k_i	Federkonstante eines „Maxwell-Slip“-Elementes
K_I	Integralverstärkung
$\mathbf{K_K}(k)$	Kalman-Verstärkung
k_{LM}	Kraftkonstante des Linearmotors
K_P	Proportionalverstärkung
$\mathbf{K_{RLS}}(k)$	Korrekturmatrix
k_V	Verstärkungsfaktor der Leistungselektronik
l_r	Lernrate im BPA
m	Masse
M	Anzahl der „Maxwell-Slip“-Elemente
$\mathbf{m}(k)$	Systemeingangsvektor
M_{ADES}	Anzahl der Tastschritte zur Zustandsschätzung
$\vec{n}, \vec{n}(s)$	(normierter) Normalenvektor

N	Anzahl der Inkremente
N_T	Ordnung der Taylorreihe
$\mathbf{o}(k)$	Messrauschen
$\mathbf{P}(k)$	Kovarianzmatrix des Modellfehlers
$\mathbf{Q}(k)$	Kovarianzmatrix des Prozessrauschens
r	Radius des Kreisbogens
$\mathbf{R}(k)$	Kovarianzmatrix des Messrauschens
$R, R(\varphi)$	Rotationsmatrix
\vec{r}_a	(normierte) Drehachse des Kreisbogens, $\vec{r}_a(t) = \begin{pmatrix} r_{a,x} \\ r_{a,y} \\ r_{s,z} \end{pmatrix}$
$r_{a,x}, r_{a,y}, r_{a,z}$	x-, y-, z-Komponente der Drehachse des Kreisbogens
$r_i(t), \Theta_i(t), g_i(t), q_i(t), p_i(t)$	Lineare Koeffizienten des DNLRX-Reibmodells
\vec{r}_s	(normierter) Richtungsvektor der Strecke, $\vec{r}_s(t) = \begin{pmatrix} r_{s,x} \\ r_{s,y} \\ r_{s,z} \end{pmatrix}$
$r_{s,x}, r_{s,y}, r_{s,z}$	x-, y-, z-Komponente des Richtungsvektors der Strecke
r_{start}, r_{ziel}	Radius des Schmiegekreises einer Blosskurve am Kurvenanfang und Kurvenende
\mathbf{S}	Vorfiltermatrix
$s, s(t)$	(Bahn-) Ortsfunktion
$\vec{s}, \vec{s}(t)$	(Kurven-) Ortsfunktion, $\vec{s}(t) = \begin{pmatrix} s_x \\ s_y \\ s_z \end{pmatrix}$
\vec{s}_{mitte}	Ortsvektor des Kreismittelpunktes, $\vec{s}_{mitte} = \begin{pmatrix} s_{mitte,x} \\ s_{mitte,y} \\ s_{mitte,z} \end{pmatrix}$
$s_{mitte,x}, s_{mitte,y}, s_{mitte,z}$	x-, y-, z-Koordinate des Kreismittelpunktes
s_{start}	(Bahn-) Startort

\vec{s}_{start}	(Kurven-) Startort, $\vec{s}_{start} = \begin{pmatrix} s_{start,x} \\ s_{start,y} \\ s_{start,z} \end{pmatrix}$
$s_{start,x}, s_{start,y}, s_{start,z}$	Startort der x-, y-, z-Achse
s_x, s_y, s_z	Ortsfunktion der x-, y-, z-Achse
s_{ziel}	(Bahn-) Zielort
\vec{s}_{ziel}	(Kurven-) Zielort, $\vec{s}_{ziel} = \begin{pmatrix} s_{ziel,x} \\ s_{ziel,y} \\ s_{ziel,z} \end{pmatrix}$
$s_{ziel,x}, s_{ziel,y}, s_{ziel,z}$	Zielort der x-, y-, z-Achse
t	Zeit
$\vec{t}, \vec{t}(s)$	(normierter) Tangentenvektor
$T_0, T_1, T_2, \text{etc.}$	Intervallbreite
T_{end}	Integrationsgrenze bzw. Fensterbreite von ADES
\mathbf{T}_r	Transformationsmatrix (Rotation)
\vec{t}_t	Transformationsvektor (Translation)
$\mathbf{U}(k)$	Eingangsvektor des künstlichen neuronalen Netzes
$\mathbf{U}(k)$	Obere Dreiecksmatrix bei der U-D-Faktorisierung
U_A, U_B	Quadratursignale
$u_{PI}(t)$	Stellgröße PI-Erweiterung
$\mathbf{u}_R(t)$	Stellsignale des Reglers
$u_S(t)$	Steuerspannung der Leistungselektronik
$\mathbf{u}_V(t)$	Stellsignale der Vorsteuerung
$\mathbf{u}_Z(t)$	Stellsignale des Störgrößenbeobachters
$u_Z(t)$	Stellgröße Zustandsregelung
v	Glättungsparameter
$\mathbf{v}(k)$	Prozessrauschen
$v, v(t)$	(Bahn-) Geschwindigkeitsfunktion

$\vec{v}, \vec{v}(t)$	(Kurven-) Geschwindigkeitsfunktion, $\vec{v}(t) = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}$
v_{max}	(Bahn-) Geschwindigkeitsrestriktion
\vec{v}_{max}	(Kurven-) Geschwindigkeitsrestriktion, $\vec{v}_{max}(t) = \begin{pmatrix} v_{max, x} \\ v_{max, y} \\ v_{max, z} \end{pmatrix}$
$v_{max, x}, v_{max, y}, v_{max, z}$	Geschwindigkeitsrestriktion der x-, y-, z-Achse
v_S	Stribeck-Geschwindigkeit
v_{start}	(Bahn-) Startgeschwindigkeit
v_x, v_y, v_z	Geschwindigkeitsfunktion der x-, y-, z-Achse
v_{ziel}	(Bahn-) Zielgeschwindigkeit
\mathbf{w}	Vektor der Gewichte des KNN
$w(t)$	Plastischer Anteil einer Bewegung
$\mathbf{w}(t)$	Sollwerte
$w_{i, q}, w_i^O$	Gewichte des künstlichen neuronalen Netzes
w_{Δ}	Gradient eines beliebigen Gewichtes in einem KNN
x	Skalierungsfaktor zur Bestimmung der Bahnrestriktion
$x(t)$	Position
$\mathbf{x}(t)$	Zustandsvektor
$\dot{x}(t)$	Geschwindigkeit
$\ddot{x}(t)$	Beschleunigung
$\mathbf{y}(t)$	Regelgrößen
$y(t)$	Regelgröße
$z(t)$	Zustandsvariable der Borstenmodelle bzw. elastischer Anteil einer Bewegung
z_{ba}	Kritische Auslenkung der Borsten
$z_{ss}(\dot{x}(t))$	Maximale Auslenkung der Borsten
α_{ADNLRX}	Schrittweite des ADNLRX-Modells
α_o	Koeffizienten der Taylorreihe
β	Gestaltparameter

γ_i	Skalierungsfaktor
Δ	Nichtlinearer Schwellwertvektor des DNLRX-Reibmodells
Δ_i	Schwellwert eines „Maxwell-Slip“-Elementes
$\delta_i(t)$	Federauslenkung eines „Maxwell-Slip“-Elementes
$\kappa, \kappa(s)$	Krümmung
$\kappa_1, \kappa_2, \kappa_3$	Laufvariablen
λ	Laserwellenlänge
λ	Skalarer Kurvenparameter des Kreisbogens
μ	Reibkoeffizient
ξ	Vergessensrate
$\Pi_{o,i}$	Gewichte des algebraischen Zustandsschätzers
ς	Gewichtung des nichtlinearen Anteils des ADNLRX-Modells
σ_0	Mikro-elastische Steifigkeit
σ_1	Mikro-elastische Dämpfung
σ_2	Geschwindigkeitsproportionaler Faktor der viskosen Reibung
σ_s	Steifigkeitskoeffizient
\mathbf{T}	Nichtlinearer Parametervektor des GMS-Modells
$\tau, \tau(s)$	Torsion
v	Anstieg der Sigmoidfunktionen
φ	Drehwinkel, skalarer Kurvenparameter des Kreisbogens
Φ	Linearer Parametervektor
φ_{ziel}	Zielwinkel
Ψ	Vektor mit linearen Modellparametern

Abkürzungen

ADES	Algebraic Derivation Estimation Scheme
ADNLRX	Adaptive Dynamic NonLinear Regression with direct application of eXcitation
AF	Aktivierungsfunktion
AFM	Atomic Force Microscopy
AMD	Advanced Micro Devices
ANSI	American National Standards Institute
BMBF	Bundesministerium für Bildung und Forschung
BPA	Backpropagation Algorithm
CAD	Computer Aided Design
CMM	Coordinate Measurement Machine
CMOS	Complementary Metal Oxide Semiconductor
DFG	Deutsche Forschungsgemeinschaft
DNLRX	Dynamic NonLinear Regression with direct application of eXcitation
DSP	Digital Signal Processor
FF	Feedforward
FIR-Filter	Filter mit endlicher Impulsantwort
GMS	Generalized-Maxwell-Slip
I++ DME	Inspection-plusplus Dimensional Measurement Equipment
IMMS	Institut für Mikroelektronik- und Mechatronik-Systeme
ITRS	International Technology Roadmap for Semiconductors
KNN	Künstliches neuronales Netz
LuGre	Lund-Grenoble
METAS	Schweizerisches Bundesamt für Metrologie

MIT	Massachusetts Institute of Technology
MKQ	Methode der kleinsten Quadrate
NMM	Nanomessmaschine
NPM	Nanopositioniermaschine
NPMM	Nanopositionier- und Nanomessmaschinen
OP	Optimierungsproblem
PC	Personal Computer
PCMCIA	Personal Computer Memory Card International Association
PTB	Physikalisch-Technische Bundesanstalt
RLS	Recursive Least Squares
RMSE	Root Mean Square Error
SFB	Sonderforschungsbereich
TCP	Tool Center Point
USB	Universal Serial Bus

Zusammenfassung

Das Ziel der vorliegenden Arbeit ist die Entwicklung einer modellbasierten Entwurfsmethodik für Trajektorienfolgeregelungen, die auf reibbehaftete Hochpräzisionskinematiken zugeschnitten sind. Gegenstand der experimentellen Untersuchungen ist hierbei ein zweiachsiges Nanopositioniersystem mit einem Arbeitsbereich von $200 \times 200 \text{ mm}^2$ und einer Positionsauflösung von $0,1 \text{ nm}$. Experimentelle Untersuchungen zeigen, dass das Systemverhalten von starken, reibungsinduzierten Nichtlinearitäten dominiert wird.

Aus regelungstechnischer Sicht liegt somit der Fokus der Modellbildung auf einer dynamischen Reibkraftmodellierung. Im Zuge der Validierung werden ein physikalisch motiviertes Reibmodell sowie ein künstliches neuronales Netz für eine spätere Verwendung in einer Trajektorienfolgeregelung ausgewählt. Weiterhin kann während des Modellbildungsprozesses eine starke Orts- und Zeitvarianz des Systemverhaltens beobachtet werden. Um dies zu kompensieren, werden die ausgewählten Reibmodelle adaptiv gestaltet. Eine weitere wesentliche Komponente einer Trajektorienfolgeregelung ist ein Bahnplanungsalgorithmus. Aufgrund der außergewöhnlichen Anforderungen an die Dynamik sowie die Gleichmäßigkeit der zu planenden Bewegungen wird im Rahmen der vorliegenden Arbeit ein neuartiger Algorithmus entworfen, welcher rein analytisch für jeden Einzelachsregler des kinematischen Systems stetig differenzierbare Sollwerttrajektorien für Position, Geschwindigkeit, Beschleunigung, Ruck sowie Ruckanstieg berechnet. Eine geplante Kurve in 3D kann sich hierbei aus einer beliebigen Anzahl der Streckenelemente Linie und Kreis zusammensetzen. Diese Grundelemente können falls notwendig mit erweiterten Bloss-Kurven rucksprungfrei verschliffen werden. Neben der Vorsteuerung und dem Bahnplanungsmodul ist eine unterlagerte Regelung Bestandteil einer Trajektorienfolgeregelung. Es werden ein klassischer PID-Regler und ein PI-Zustandsregler mit Führungsgrößenaufschaltung entworfen. Zur weiteren Steigerung der Dynamik des geregelten Systems wurde die Trajektorienfolgeregelung um einen Störbeobachter erweitert. Dieser basiert auf dem Schätzalgorithmus von Kalman und berechnet den Einfluss unmodellierbarer Systemgrößen wie beispielsweise Temperaturschwankungen.

Nach dem Entwurf der Regelungskomponenten wird im experimentellen Teil der Arbeit die Leistungsfähigkeit der Module untersucht. Während dieser Analyse zeigt sich, dass eine Kombination aus adaptiven physikalisch motiviertem Reibmodell, Zustandsregler und Störbeobachter einen optimalen Kompromiss aus Regelgüte und Rechenzeit darstellt. Im Vergleich zu klassisch motivierten Regelungen kann die dynamische Regelabweichung im gesamten Arbeitsbereich des Demonstrators signifikant verkleinert werden. Nach diesen außerordentlich positiven Ergebnissen wird das Konzept der Trajektorienfolgeregelung auch auf die zweite Bewegungsachse übertragen und ein dezentraler Regelungsansatz umgesetzt.

Abstract

The goal of the given thesis is the development of a model-based design method for trajectory tracking control, which is focused on friction- afflicted high precision kinematics. The subject of the experimental investigations is a two-axis nano-positioning system with a working area of $200 \times 200 \text{ mm}^2$ and a resolution of 0.1 nm . Experimental research shows that the system performance is dominated by strong, friction-induced nonlinearities.

From control-technical view, the focus of the concept lies on a dynamic friction force modelling. In the course of validation, a physically motivated friction model as well as an artificial neural network is selected for a later use in a trajectory tracking controller. Furthermore, a strong space and time variance of the system behaviour can be observed during the modelling process. In order to compensate for this, the selected friction models are adaptively arranged. Another substantial component of a trajectory tracking controller is a path planning algorithm. Due to the extraordinary requirements on the dynamics as well as the uniformity of the planned movements, a novel algorithm is designed. This method computes analytically for every axis controller of the kinematic system constantly differentiable set-point trajectories for position, speed, acceleration, jerk and derivative of jerk. A planned curve in 3D can consist of any number of path elements (line and circle). When necessary, these basic elements can be smoothed with extended Bloss-curves. Other than the servo control and the path planning module, a feedback controller is a component of a trajectory tracking controller. A classical PID controller and a PI state space controller with additional reference variable input are designed. For the further improvement of the dynamics of the controlled system, the trajectory tracking controller was expanded using a disturbance observer. This observer is based on Kalman's estimation algorithm and calculates the influence of system parameters that cannot be modelled, such as for example temperature fluctuations.

After the design of control components, the efficiency of the modules is examined in the experimental part of the work. During this analysis it appears that an optimal compromise between performance and computing time is obtained by combining together an adaptive, physically motivated friction model, the designed state-space controller and disturbance observer. In contrast with the classically motivated controls, the dynamical offset can be made significantly smaller in the whole working space. After such excellent results, the concept of trajectory tracking control is also implemented on the second motion axis of the experimental set-up and a decentralized control approach is realized.

1 Einleitung

Seit der Mitte des vergangenen Jahrhunderts ist die ununterbrochene Miniaturisierung in vielen Bereichen der Technik ein Innovationsmotor. Besonders die Halbleiterindustrie ist geprägt von dem Wunsch nach zunehmender Rechenleistung, höherer Schaltgeschwindigkeit und geringerem Strombedarf bei gleichzeitiger Steigerung der Integrationsdichte [14]. Bereits 1965 formulierte Gordon Moore das Mooresche Gesetz, welches für die Halbleiterindustrie nach wie vor den Rahmen ihrer Entwicklungsziele vorgibt. Es sagt nach der heute vorherrschenden Auslegung aus, dass sich die Anzahl der Transistoren auf einem Computerchip alle 24 Monate verdoppelt [59]. Moderne Haupt- und Grafikprozessoren werden derzeit in Strukturgrößen von 45 bis 65 nm gefertigt, eine weitere Verkleinerung auf 32 nm ist bereits in Vorbereitung und soll noch 2009 Serienreife erlangen. Diese Entwicklung wird sich laut der „International Technology Roadmap for Semiconductors“ (ITRS) weiter fortsetzen, und so wird für 2020 ein $\frac{1}{2}$ Pitch-Wert von 14 nm prognostiziert [60]. Zeitgleich hat aus betriebswirtschaftlichen Gründen die Fläche, auf der diese Strukturen während der Produktion angeordnet werden, permanent zugenommen. Auch diese Feststellung kann exemplarisch anhand der Halbleiterfertigung belegt werden. Nach derzeitigem Entwicklungsstand werden Wafer mit einem Durchmesser von 12 Zoll (300 mm) verarbeitet, und innerhalb der nächsten drei Jahre wird eine Steigerung auf 18 Zoll (450 mm) angestrebt [58].

Neben dem angeführten Beispiel der CMOS-Industrie müssen auch andere Technologiebereiche, in denen Strukturen kleiner 100 nm eine entscheidende Rolle spielen, erwähnt werden. Hierzu zählen beispielsweise die Wissenschaftsfelder der Chemie, der Optik und natürlich die Lebenswissenschaften („Life Sciences“) [84]. Anwendung fand die Nanotechnologie bis dato in Produkten der Pharmazie, der Elektro- und Halbleitertechnik, des Automobilbaus, des Maschinenbaus, der Umwelttechnik sowie der chemischen Industrie. Laut der Einschätzung einer durch das Bundesministerium für Bildung und Forschung (BMBF) beauftragten Studie sind bereits heute in Deutschland mindestens 50.000 Arbeitsplätze direkt oder indirekt im Bereich der Nanotechnologie anzusiedeln. Bis zum Jahr 2015 wird eine Ausdehnung dieser Schlüsseltechnologie auf nahezu alle Industriebereiche erwartet [27].

Ermöglicht wurde diese Entwicklung durch außerordentliche Fortschritte auf dem Gebiet der Nanomesstechnik in den vergangenen 25 Jahren. Erst seit der Erfindung des Rastertunnel- [25], des Rasterkraft- [24] oder des Rastersondenmikroskopes [133] können Strukturen mit einer Auflösung im Nanometerbereich gemessen, visualisiert und gegebenenfalls manipuliert werden. Aufgrund der regen Forschungstätigkeit in der Nanomesstechnik wurden bis dato eine Vielzahl von Rastertunnel-, Rasterkraft- und Rastersondenmikroskope zur Marktreife gebracht. Die kommerziell verfügbaren Geräte sind hochgradig spezialisierte Messapparaturen, die lediglich ein sehr kleines Messvolumen ($>1 \text{ mm}^3$) abdecken [16, 17, 128]. Diese Ge-

räte eignen sich hervorragend für eine Vielzahl von Anwendungen in den genannten Technologiefeldern und sind mittlerweile Standard in gut ausgestatteten Forschungslabors [36].

Wie aber bereits eingangs erwähnt wurde, erfordern aktuelle Entwicklungen in nahezu allen Bereichen der Nanotechnologie zunehmend größere Messbereiche bei gleichbleibender Auflösung und Reproduzierbarkeit. Somit begrenzen die derzeit verfügbaren Messvolumina die Anwendbarkeit der existenten Messgeräte und infolgedessen arbeiten weltweit mehrere Forschungsgruppen am Bau von so genannten Nanopositionier- und Nanomessmaschinen (NPMM). Diese Maschinen verbinden Auflösungen im Subnanometerbereich mit makroskopischen Messvolumina, was enorme Anforderungen an das Design solcher Geräte stellt. Die größte Herausforderung stellt hierbei die Realisierung einer Nanometerskala im gesamten Messvolumen dar und es müssen sowohl die Reproduzierbarkeit als auch die Rückführbarkeit gewährleistet bleiben [46].

1.1 Die Nanomessmaschine 1

Das Institut für Prozessmess- und Sensortechnik der Technischen Universität Ilmenau erkannte frühzeitig diesen Bedarf und entwickelte mit dem Institut für Technische Informatik und Ingenieurinformatik, dem Zentrum für Bild- und Signalverarbeitung e. V. sowie der SIOS Meßtechnik GmbH im Rahmen eines Thüringer Verbundprojektes die in Abbildung 1.1 dargestellte Nanopositionier- und Nanomessmaschine (NPMM).

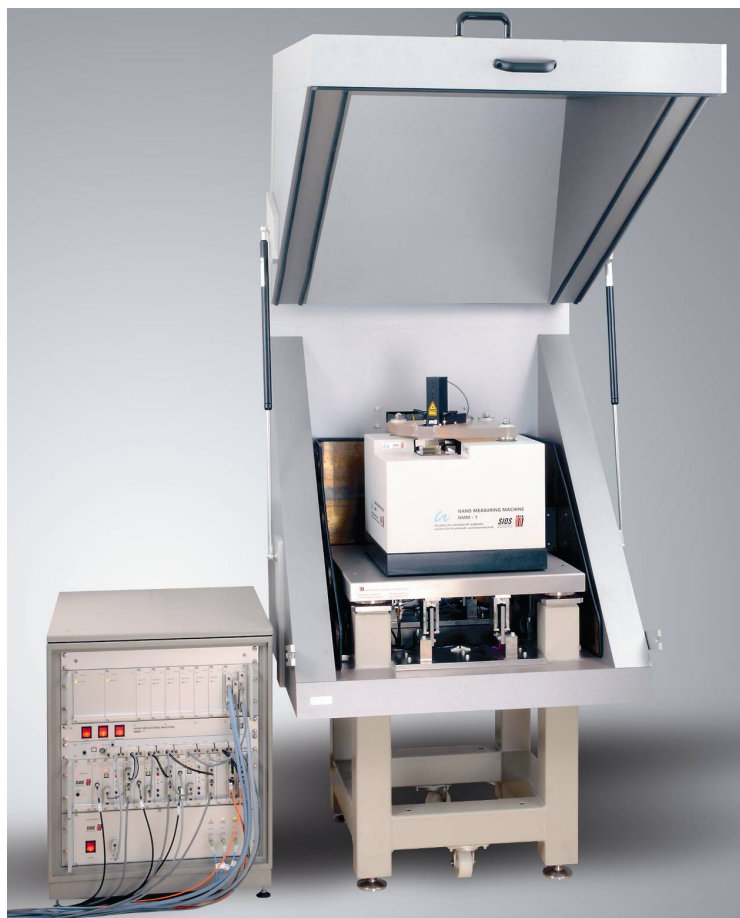


Abbildung 1.1: Nanomessmaschine 1 (NMM-1)

Das Konzept dieser zwischen 1996 und 2000 entwickelten Maschine ist derart wegweisend, dass sie unter dem Namen Nanomessmaschine 1 (NMM-1) zur Marktreife gebracht werden konnte und nunmehr weltweit eingesetzt wird [116]. Der Bewegungsbereich der NMM-1 beträgt 25 mm in x- und y-Richtung sowie 5 mm in z-Richtung. Insbesondere der große Messbereich verbunden mit der erreichten Auflösung von 0,1 nm und der Reproduzierbarkeit von unter zehn Nanometern werden derzeit als weltweite Alleinstellungsmerkmale auf dem Gebiet der 3D-Nanopositionier- und Nanomesstechnik gewertet [49, 63, 64, 122]. Des Weiteren bietet die NMM-1 eine ausgesprochen große Flexibilität, denn es kann eine Vielzahl von optischen und taktilen Antastsystemen appliziert werden.

Die Präzision und Reproduzierbarkeit der NMM-1 ist nur erreichbar, weil im Gegensatz zu konventionellen Koordinatenmessmaschinen das Abbe'sche Komparatorprinzip [1] in allen Messachsen und an allen Messpositionen eingehalten wird [49, 110]. Somit können Messfehler erster Ordnung prinzipbedingt minimiert werden, denn das Messobjekt und die Messeinrichtung sind fluchtend zueinander angeordnet. In Abbildung 1.2 ist das in der NMM-1 realisierte Prinzip schematisch dargestellt. Die zu vermessende Probe ruht auf einer Spiegelecke aus Zerodur® [112]. Die drei zur Längenmessung dienenden Interferometer des Typs SP-500 [115] sind derart ausgerichtet, dass sich die Laserstrahlen im Antastpunkt schneiden. Darüber hinaus werden Verkippungen der Spiegelecke um die X- und Y-Achse durch zwei hochauflösende Winkelsensoren, die nach dem Autokollimationsprinzip arbeiten [139], erfasst.

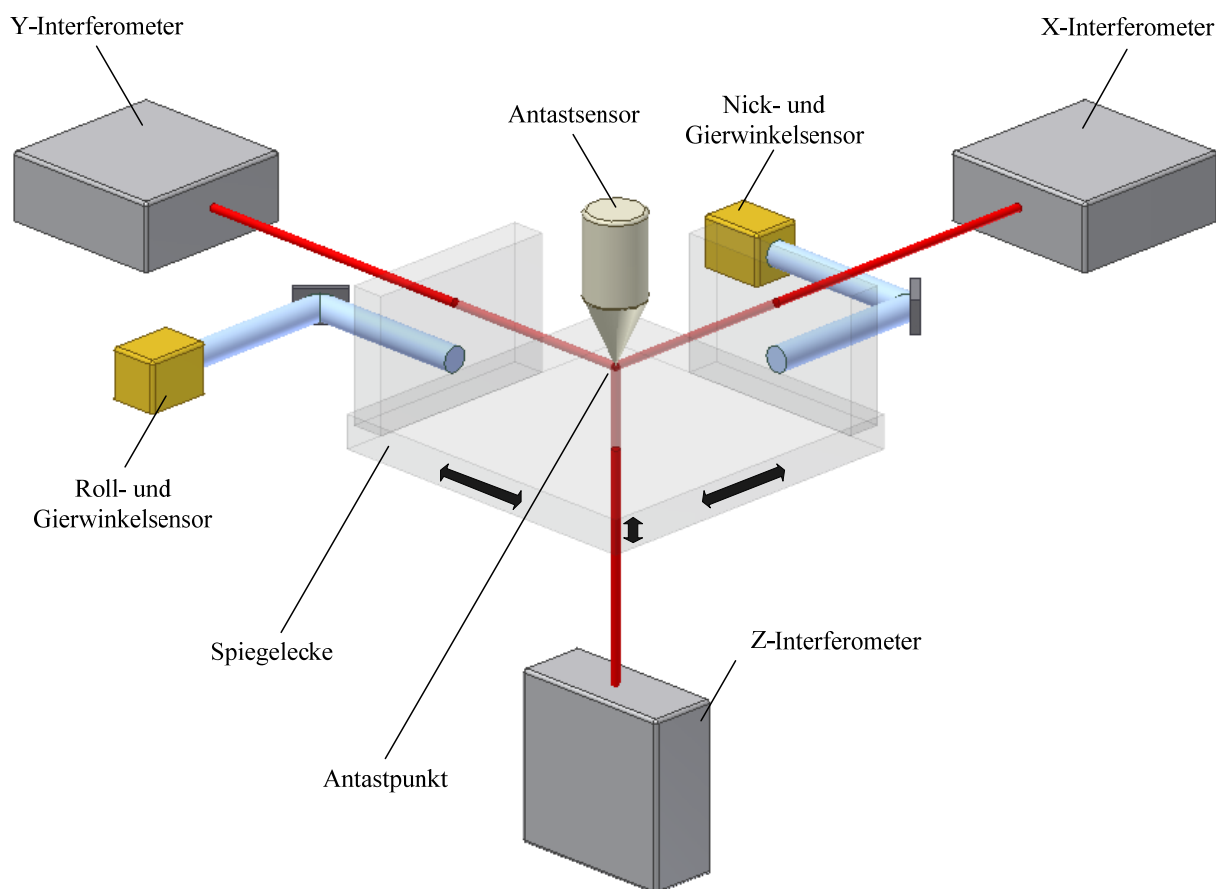


Abbildung 1.2: Visualisierung des Abbe'schen Komparatorprinzips in allen drei Messachsen

Werden diese Rotationen permanent ausgegeregelt, kann das positionsfeste Tastsystem an jeder beliebigen Position der Spiegelecke als Nullpunktindikator angesehen werden.

Zur Verdeutlichung zeigt Abbildung 1.3 eine schematische Darstellung des konstruktiven Aufbaues der NMM-1. Auf einem Fundament aus Granit ruht ein metrologischer Rahmen aus Zerodur®, welcher alle Längen- und Winkelmesssysteme aufnimmt. Dadurch wird gewährleistet, dass alle Messsysteme eine feste Anordnung bilden. Dies ist notwendig, damit die Einhaltung des Abbe'schen Komparatorprinzips gegenüber thermischen und mechanischen Störeinflüssen gewährleistet werden kann. Die Spiegelecke, welche das Messobjekt trägt, ist auf einem dreiachsigen Führungssystem befestigt. Die zwei horizontalen Führungen des Typs MT-120 der Firma MICOS GmbH [86] sind um 90 Grad versetzt aufeinander montiert und bilden die X- und Y-Achse. Angetrieben werden diese beiden Bewegungsachsen von homopolaren Linearantrieben des Typs LA 13-30-000 der Firma BEI-KIMCO [21]. Die Z-Achse wird mithilfe von drei zylindrischen Kugelführungen gelagert und der Antrieb erfolgt durch vier symmetrisch angeordnete Linearmotoren des Typs LA 13-12-000 [49]. Auf der Z-Plattform ist letztendlich die Spiegelecke angebracht. Sie besteht aus einer Grundplatte und zwei seitlich angespannten Platten. Die Unterseite der Grundplatte sowie die Außenseiten der angespannten Platten sind in höchster Ebenheit gefertigt und verspiegelt.

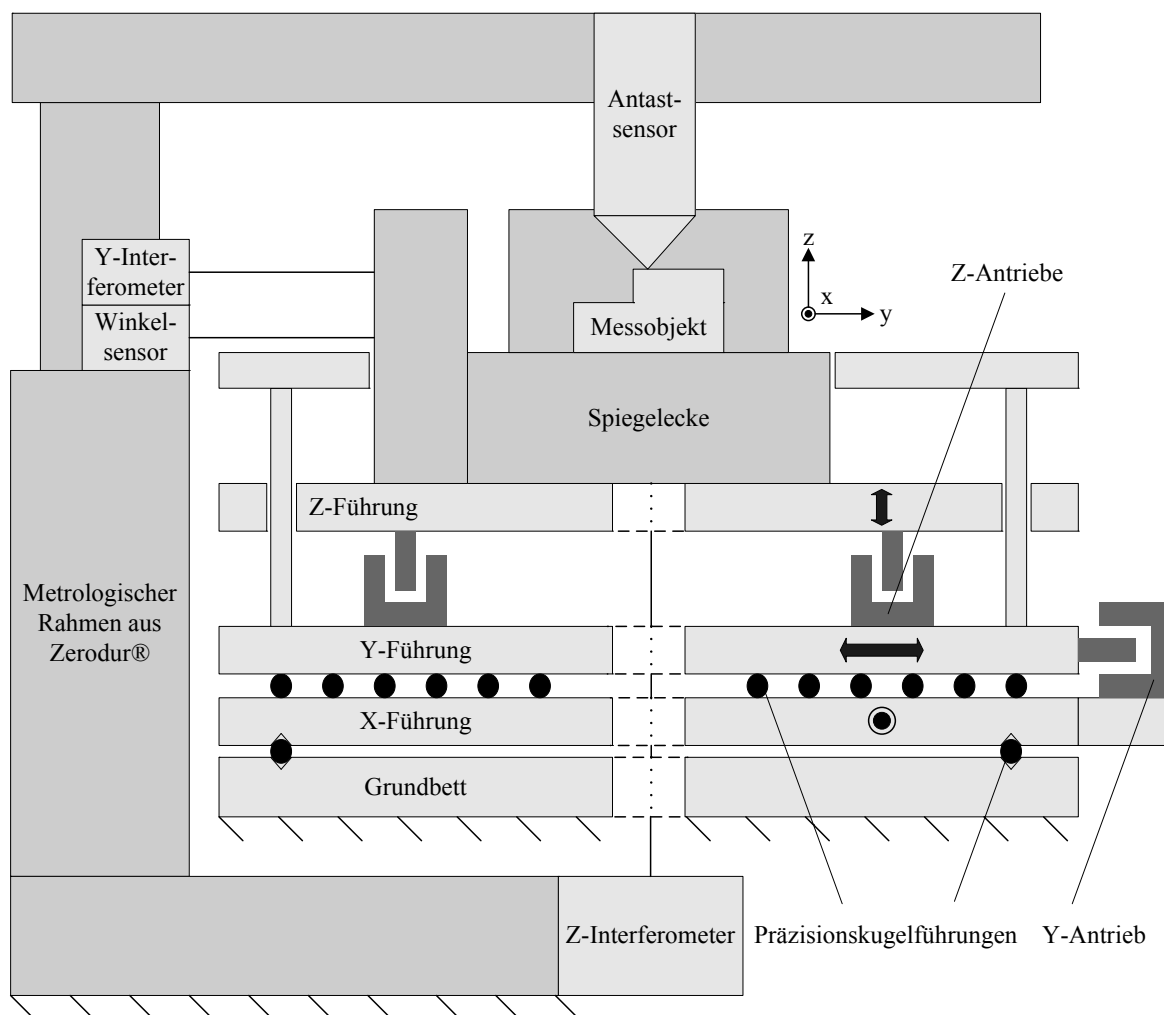


Abbildung 1.3: Schematische Darstellung des Grundaufbaues der NMM-1

Dies ist unerlässlich, denn sie dienen als Reflektionsflächen für die Längen- und Winkelmesssysteme und haben so direkten Einfluss auf die Güte der Messergebnisse [110]. Um systematische oder zufällige Führungsfehlerabweichungen in allen Bewegungsachsen zu kompensieren, werden alle fünf messtechnisch erfassten Freiheitsgrade geregelt. Lediglich Rotationen um die Z-Achse werden nicht ausgeregelt, was aber laut [110] keinen signifikanten Einfluss auf die Messunsicherheit der NMM-1 hat.

Tabelle 1.1: Entwicklungen auf dem Gebiet der Nanopositionier- und Nanomesstechnik

Bezeichnung	Entwickler/Hersteller	Leistungsdaten
NMM-1 [49]	TU Ilmenau, SIOS Meßtechnik GmbH	Messvolumen: 25x25x5 mm ³ Reproduzierbarkeit: < 10 nm Auflösung: 0,1 nm
Small CMM [77, 92]	National Physical Laboratory, UK	Messvolumen: 50x50x50 mm ³ Unsicherheit: < 50 nm Auflösung: 3 nm
High Precison 3D-CMM bzw. „F25“ [28]	TU Eindhoven bzw. Carl Zeiss Industrielle Messtechnik GmbH	Messvolumen: 140x140x100 mm ³ Unsicherheit: < 250 nm Auflösung: 0,25 nm
NanoCMM1 [127]	TU Eindhoven	Messvolumen: 50x50x4 mm ³ Unsicherheit: 25 nm Auflösung: 1 nm
μCMM [26]	Physikalisch-Technische Bundesanstalt (PTB)	Messvolumen: 24x34x35 mm ³ Unsicherheit: < 100 nm
Ultraprecision CMM [69]	Schweizer Bundesamt für Metrologie (METAS)	Messvolumen: 90x90x38 mm ³ Messunsicherheit: < 80 nm Auflösung: 3 nm
Sub-Atomic Measuring Machine [51, 52]	University of North Carolina at Charlotte, Massachusetts Institute of Technology (MIT)	Messvolumen: 25x25x0,1 mm ³ Angestrebte Auflösung: 0.01 nm
micro-CMM [42]	National Taiwan University	Messvolumen: 25x25x10 mm ³ Reproduzierbarkeit: 30 nm Auflösung: 10 nm
Ultrahigh Accurate 3-D Profilometer [121]	Panasonic	Messvolumen: 400x400x90 mm ³ Reproduzierbarkeit: 50 nm Messfehler: 10 nm

Wie in Abbildung 1.1 zu sehen ist, besteht die NMM-1 aus einer weiteren Komponente, der so genannten Versorgungs- und Auswerteeinheit (19'' Gehäuse). In dieser sind die Laserrohre der Längen- und Winkelmesssysteme, die Leistungselektronik der Antriebe, die Messwertverarbeitung sowie das Steuerungs- und Regelungssystem untergebracht. Auch diese räumliche Trennung der Baugruppen trägt entscheidend zur Qualität der Messergebnisse bei, da sämtliche Komponenten, die Wärme erzeugen, außerhalb des eigentlichen Messsystems platziert sind. Als Nutzerschnittstelle dient die Software Matlab/Simulink® [123] auf einem Host-PC. Die vom Operator nach dem I++/DME Standard [55] spezifizierten Messbefehle werden über eine USB Schnittstelle an das DSP-basierte Regelungs- und Steuerungssystem übergeben und anschließend ausgeführt.

Diese Beschreibung der NMM-1 gibt natürlich nur eine kurze Zusammenfassung der Konstruktionsprinzipien sowie der grundsätzlichen Funktionsweise. Für tiefer gehende Informationen bezüglich NMM-1 sei an dieser Stelle auf [49] verwiesen.

Zur besseren Einordnung der NMM-1 soll im Folgenden kurz auf weitere Entwicklungen im Bezug auf Nanopositionier- und Nanomessmaschinen eingegangen werden. In Tabelle 1.1 werden ausgewählte internationale Forschungsprojekte dargestellt, die sich zum Ziel gesetzt haben, 3D- Koordinatenmessmaschinen mit Reproduzierbarkeiten im zweistelligen Nanometerbereich zu entwickeln. In diesem Zusammenhang sollte aber erwähnt werden, dass es sich bisher bei den vorgestellten Maschinen lediglich um Prototypen handelt.

1.2 Zielstellung der Arbeit

Infolge des wissenschaftlichen und wirtschaftlichen Erfolges der NMM-1 wurde im Jahr 2002 durch die Deutsche Forschungsgemeinschaft (DFG) der Sonderforschungsbereich (SFB) 622 „Nanopositionier- und Nanomessmaschinen“ an der Technischen Universität Ilmenau eingerichtet. Die Zielstellung des SFB 622 ist die Erarbeitung der wissenschaftlichen und technischen Grundlagen für NPM-Systeme mit einem Messvolumen von bis zu $450 \times 450 \times 80 \text{ mm}^3$. Darüber hinaus werden neuartige Antast- und Bearbeitungssysteme in diesem SFB entwickelt und in NPM-Systeme eingebunden. Am SFB 622 sind 13 Fachgebiete sowie das Institut für Mikroelektronik- und Mechatronik-Systeme (IMMS) gGmbH [57] beteiligt. Die definierten Ziele stellen natürlich außerordentliche Anforderungen an Genauigkeit, Reproduzierbarkeit und vor allem an die Positioniergeschwindigkeit der zu entwickelnden mechatronischen Systeme. Somit muss auch das Regelungssystem zur Führung eines NPM-Systems erheblich mehr leisten als eine herkömmliche Positionsregelung. Um mit einer NPM-Maschine (NPM) beispielsweise AFM¹-Messungen durchzuführen, sind aus Sicht der Messtechnik Bahngeschwindigkeiten von 10 bis 40 $\mu\text{m/s}$ bei gleichzeitiger Einhaltung von Positionsunsicherheiten $< 1 \text{ nm}$ wünschenswert. Ähnliche Anforderungen an das Regelungs- und Steuerungssystem ergeben sich auch beim optischen Scannen mithilfe einer NPM. Hier werden

¹ Atomic Force Microscopy

Bahngeschwindigkeiten von maximal 30 mm/s verbunden mit Positionsunsicherheiten unter 1 μm gefordert.

In der Nanopositionier- und Nanomesstechnik dominieren bislang klassische, experimentell motivierte Regelungsansätze [48, 49, 64, 105, 125]. Nachteilig an einer solchen Herangehensweise sind sowohl der hohe experimentelle Aufwand bei der Reglerbemessung als auch die fehlende Allgemeingültigkeit des entworfenen Reglers sowie die stark beschränkte Messgeschwindigkeit. Darüber hinaus sind diese Regelungskonzepte nicht in der Lage, Schleppfehler, die während einer dynamischen Bewegung entstehen, zu kompensieren. NPM-Systeme mit einem Arbeitsbereich von bis zu $450 \times 450 \times 80 \text{ mm}^3$ bei einer Auflösung von 0,1 nm erfordern aber angesichts des sehr großen Messvolumens eine deutlich gesteigerte Positionierdynamik, um realistische Messzeiten zu ermöglichen.

Eine Lösung dieses Problems stellen modellbasierte Regelungsmethoden dar, die die erzielbare Positionierdynamik erheblich steigern können. Eine Gruppe solcher Methoden sind die aus der Robotik stammenden Trajektorienfolgeregelungen, welche eine exzellente Bahntreue während der Bewegung gewährleisten [31, 34, 35, 91]. Besonders im Hinblick auf den Einsatz von 3D-Microtastern ist dies eine essentielle Voraussetzung, um Beschädigungen an der Messsensorik während hochdynamischer Positioniervorgänge zu vermeiden.

Bislang fehlen auf dem Gebiet der Nanopositionier- und Nanomesstechnik sowohl die geeigneten Modellansätze als auch eine spezifisch angepasste Methodik aus der Regelungstheorie. Das Ziel der vorliegenden Arbeit ist daher die Entwicklung einer Entwurfsmethodik für Trajektorienfolgeregelungen, die in NPM-Systemen eingesetzt werden können.

1.3 Aufbau der Arbeit

Alle Ergebnisse basieren auf experimentellen Untersuchungen, die an einem zweiachsigen Demonstrator mit einem Bewegungsbereich von $200 \times 200 \text{ mm}^2$ durchgeführt wurden. Der mechanische und elektronische Aufbau dieses mechatronischen Systems wird ausführlich im zweiten Kapitel dargestellt. Darauf aufbauend findet im dritten Kapitel die Modellierung des dynamischen Systemverhaltens statt. Hierbei stellt sich heraus, dass die Reibung, welche durch die Kugelführungen in das System eingetragen wird, im Nanometerbereich einen dominanten Einfluss auf die Systemdynamik ausübt. Diese starke Nichtlinearität, welche im makroskopischen Bereich leicht kompensiert werden könnte, verursacht bei einer Positionierung mit Nanometerpräzision erhebliche Störungen und muss daher ausgesprochen präzise mit in die Modellbeschreibung einbezogen werden. Um dies zu erreichen, werden verschiedene dynamische Reibmodelle identifiziert und deren Leistungsfähigkeit auf Nanometerebene analysiert. Nach der erfolgreichen Modellbildung folgt im vierten Kapitel dieser Arbeit die Vorstellung des modellbasierten Regelungskonzeptes. Dieses besteht aus einem Bahnplanungsalgorithmus vierter Ordnung, einer unterlagerten Regelung, einer adaptiven Vorsteuerung sowie einer Störbeobachtung. Dieses modellbasierte Regelungssystem wird im fünften Kapitel für eine Achse des Demonstrators erprobt und dessen Leistungsfähigkeit im Vergleich zu dem bisher eingesetzten Konzept bewertet. Weiterhin wird in diesem Kapitel der Einfluss

der einzelnen Regelungskomponenten auf die Güte des Gesamtsystems erörtert. Auf Basis dieser Erkenntnisse wird im darauffolgenden Kapitel eine modellbasierte dezentrale Regelung für beide Achsen des Demonstrators vorgestellt und deren Leistungsfähigkeit gezeigt. Die vorliegende Arbeit schließt mit einer Zusammenfassung sowie einem Ausblick auf weiterführende Forschungsthemen im Zusammenhang mit Regelungssystemen für Nanopositionier- und Nanomessmaschinen ab.

2 Experimenteller Aufbau

In der vorliegenden Arbeit wird ein zweiachsiges Nanopositioniersystem betrachtet, das in Abbildung 2.1 dargestellt ist. Es ist zu erkennen, dass der Demonstrator auf einem optischen Breadboard befestigt ist, welches wiederum auf einem schwingungsgedämpften Fundament ruht. Das Nanopositioniersystem besitzt einen Bewegungsbereich von $200 \times 200 \text{ mm}^2$ und verfügt wie die NMM-1 über ein interferenzoptisches Messsystem mit einer Positionsauflösung von $0,1 \text{ nm}$ [115]. Zum Schutz vor Verschmutzungen und zur Dämpfung von Schallwellen ist der gesamte Aufbau einschließlich dem Breadboard von einem Gehäuse aus Styrodur® C [19] und aixFOAM® [109] umgeben.

Prinzipiell folgt auch dieser xy-Demonstrator den Konstruktionsprinzipien der NMM-1 und lässt sich als mechatronisches System in die Komponenten Sensorik, Mechanik und Aktorik unterteilen. Nach dieser Systematik gliedert sich auch das vorliegende Kapitel, wobei neben den schon erwähnten Baugruppen auch das Positionsreferenzsystem sowie das zur Steuerung und Regelung eingesetzte Echtzeitsystem in einem eigenen Abschnitt näher erläutert werden.

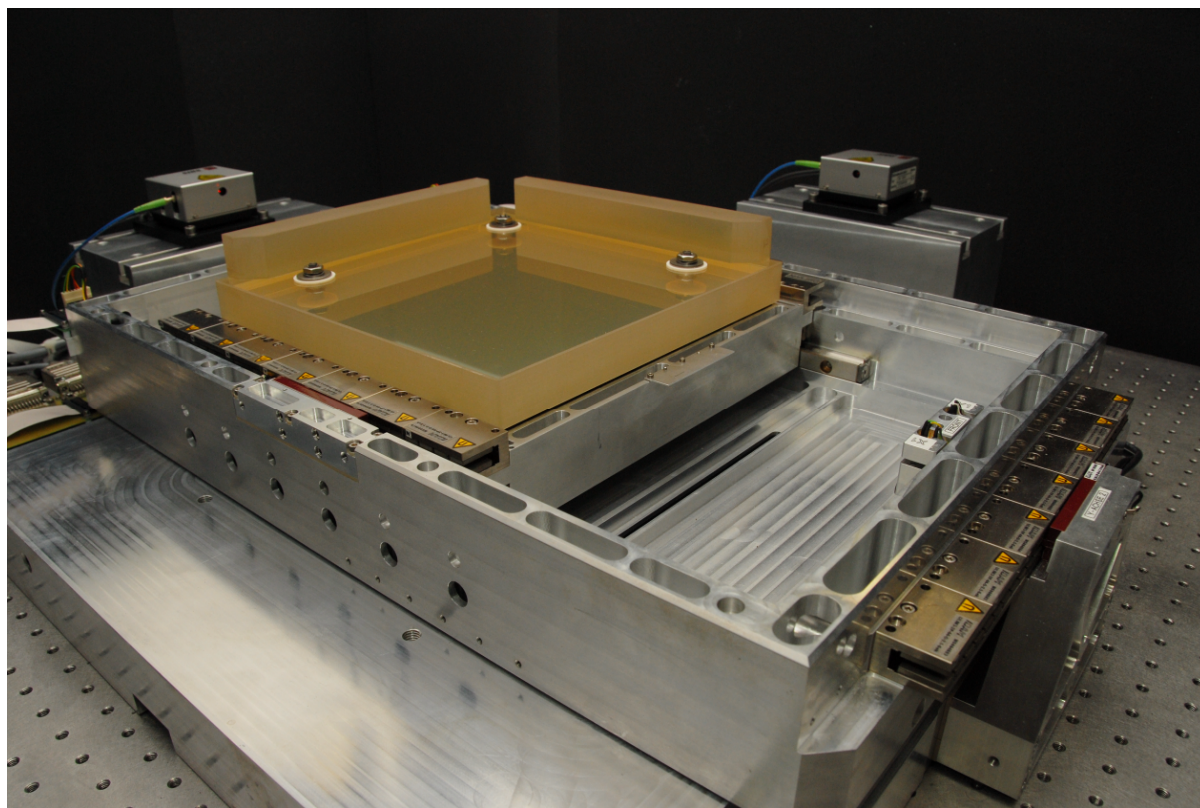
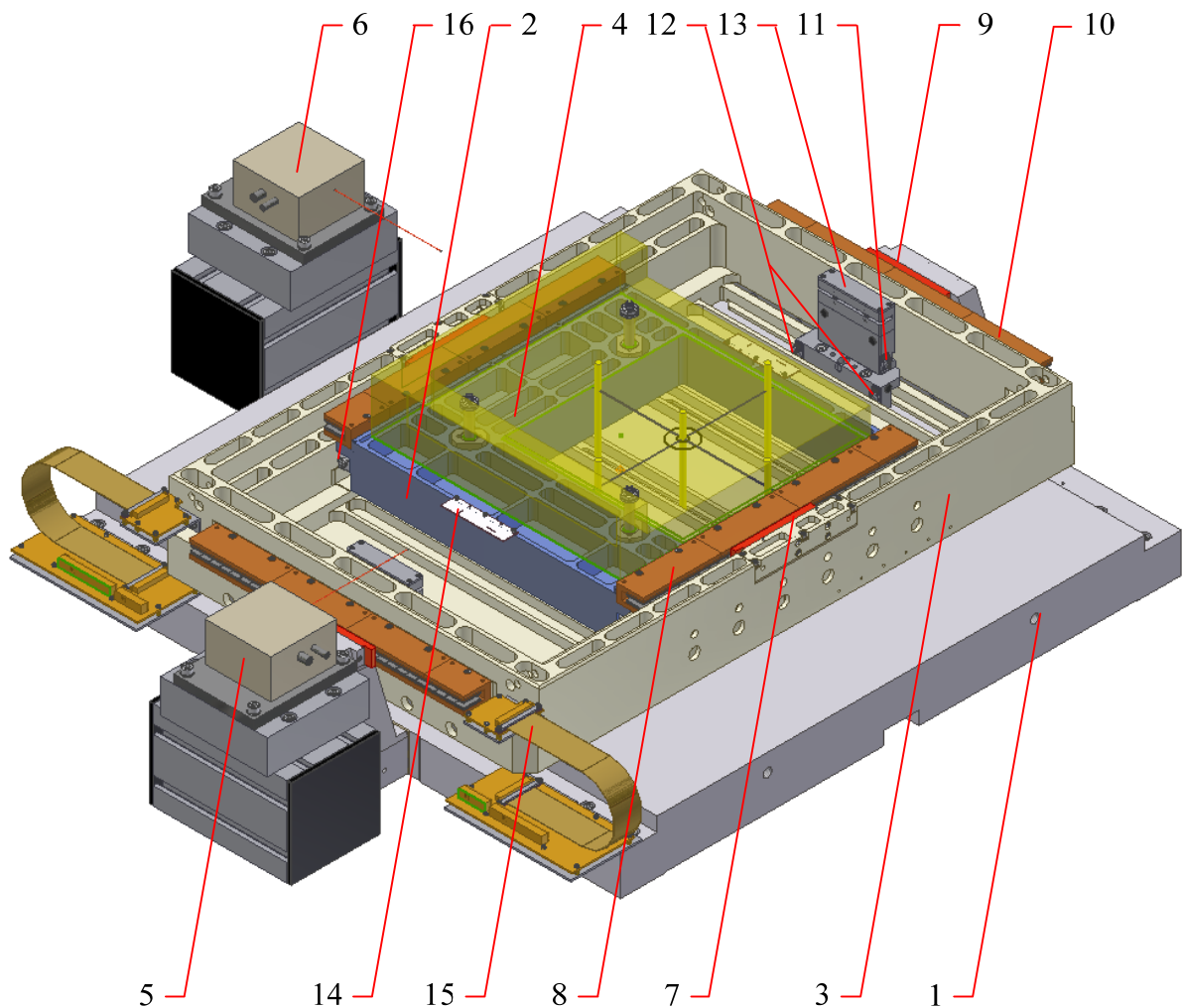


Abbildung 2.1: Zweiachsiges Nanopositioniersystem

Um einen Überblick über die mechanischen sowie elektronischen Komponenten des Nanopositioniersystems zu geben, ist in Abbildung 2.2 ein CAD-Modell des beschriebenen Versuchsaufbaus zu sehen.



- | | |
|------------------------------------|-------------------------------------|
| 1 Grundplatte | 9 Primärteil Linearmotor Y-Achse |
| 2 Slider X-Achse | 10 Sekundärteil Linearmotor Y-Achse |
| 3 Slider Y-Achse | 11 Endlagenschalter X-Achse |
| 4 Präzisionsspiegelecke | 12 Endlagenschalter Y-Achse |
| 5 Messkopf Interferometer X-Achse | 13 Positionsreferenzsystem Gestell |
| 6 Messkopf Interferometer Y-Achse | 14 Positionsreferenzsystem Fahne |
| 7 Primärteil Linearmotor X-Achse | 15 Flexbandkabel |
| 8 Sekundärteil Linearmotor X-Achse | 16 Kugelführung X-Achse |

Abbildung 2.2: CAD-Modell des verwendeten NPM-Systems mit Erläuterung der wesentlichen Komponenten

2.1 Mechanischer Aufbau

Die mechanischen Komponenten des zweiachsigen Demonstrators wurden im Rahmen des Sonderforschungsbereichs 622 am Institut für Mikroelektronik- und Mechatronik-Systeme gGmbH konstruiert [57]. Im Anschluss daran wurde das Nanopositioniersystem von der SIOS Messtechnik GmbH gefertigt und aufgebaut. Wie in Abbildung 2.3 zu sehen ist, sind auf einer Grundplatte zwei Slider mit orthogonalen Bewegungsrichtungen ineinander verschachtelt montiert. Der Y-Slider ist über zwei parallele Präzisionskugelführungen vom Typ R6-300-RF-SQ-HA der Firma Schneeberger AG mit der Grundplatte verbunden [111]. Auf die gleiche Art und Weise ist der X-Slider in dem Y-Slider gelagert. Dieses Konstruktionsprinzip wird als Doppel-H-Struktur bezeichnet, denn das Führungssystem einer Achse besitzt, wie in Abbildung 2.3 gut zu erkennen ist, die Form eines H's. Abbildung 2.4 zeigt eine Detailansicht der verwendeten Kugelführungen. Zwei dieser Führungen bilden ein Lager. Zwischen ihnen befindet sich ein Kugelkäfig des Typs AK-RF mit 15 kugelförmigen Umlaufkörpern des Typs SK. Geschmiert werden die Linearführungen mit einem Wälzlagerfett auf Lithiumseifenbasis nach DIN 51502.

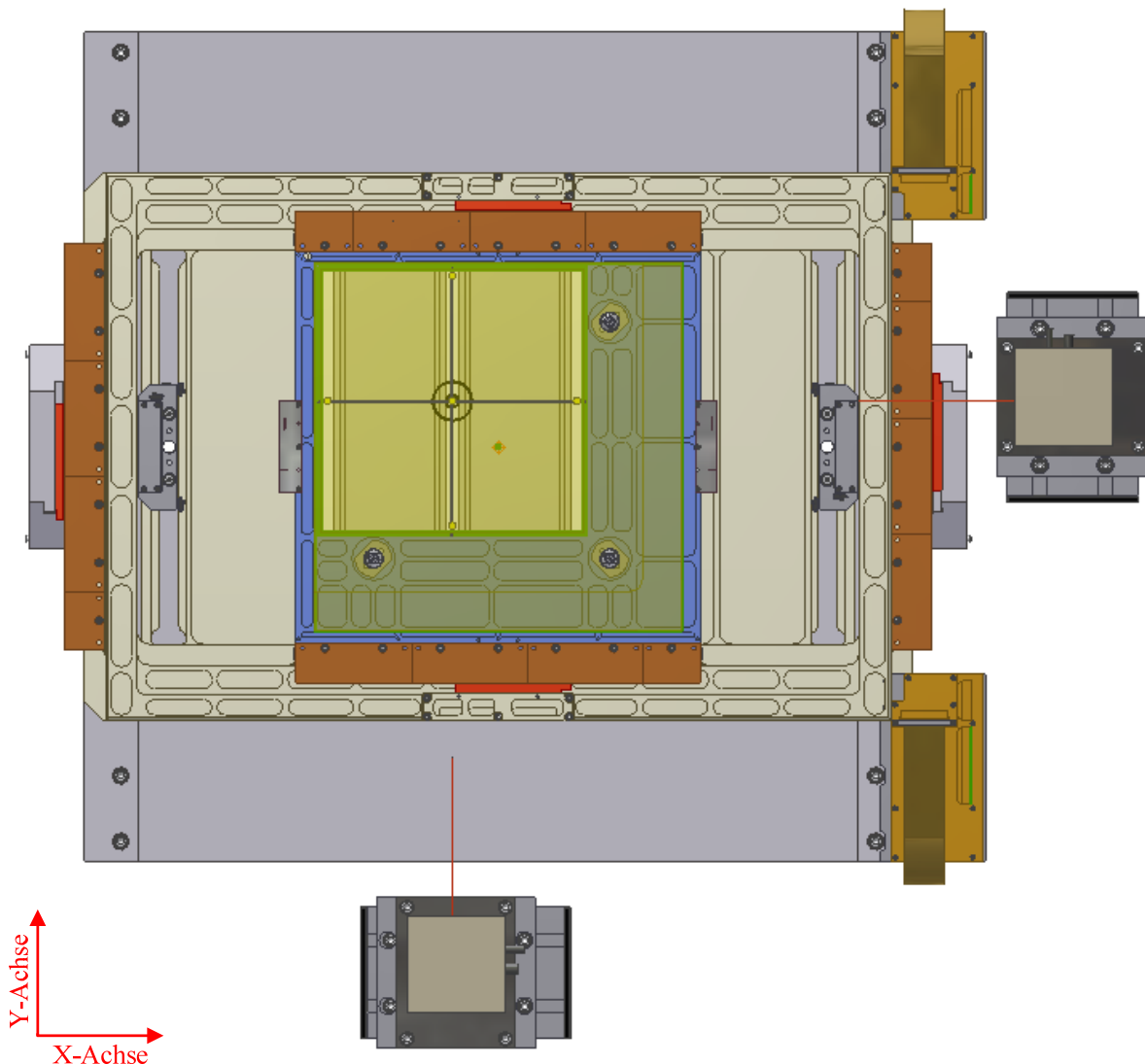


Abbildung 2.3: Draufsicht auf den zweiachsigen Demonstrator

Besonders im Hinblick auf die Vakuumtauglichkeit von NPM-Systemen mit großen Bewegungsbereichen ist diese Art der Lagerung äußerst zweckmäßig. Darüber hinaus konnte in [49] gezeigt werden, dass, besonders in der Nanopositioniertechnik, Wälzführungen alternativen Lagerungsprinzipien überlegen sind. Aus regelungstechnischer Sicht besitzen diese Führungen jedoch den entscheidenden Nachteil, dass sie ein stark nichtlineares Verhalten, verursacht durch Reibung, in das mechanische System eintragen (siehe Abschnitt 3.4). Jede der zwei Bewegungsachsen wird von zwei Linearmotoren angetrieben (siehe Abschnitt 2.3). Da auch die Motoren der X-Achse mit Strom versorgt werden müssen, verbinden zwei Flexbandkabel die Y-Achse mit Leiterkarten, die auf der Grundplatte montiert sind. Nahezu alle mechanischen Teile des Positioniersystems sind aus der Aluminiumlegierung EN AW 5083 gefertigt. Auf der X-Achse befindet sich eine Auflagefläche, auf der die zu untersuchenden Proben platziert werden können. Dieser Probenstisch stellt gleichzeitig auch die Spiegelecke dar, die das Laserlicht der beiden Interferometer (siehe Abschnitt 2.2) senkrecht reflektiert. Da die Spiegelecke höchsten messtechnischen Ansprüchen genügen muss [49], besteht sie aus Zerodur® und wurde von der Carl Zeiss Jena GmbH gefertigt. Um aus der relativen die absolute Position ermitteln zu können, muss die Spiegelecke und somit der Probenstisch zuvor mithilfe eines Positionsreferenzsystems in einen Initialzustand mit bekanntem Bezugspunkt gebracht werden (siehe Abschnitt 2.4). Des Weiteren sind sechs Endlagenschalter an den Sockeln der zwei Positionsreferenzsysteme angebracht. Zwei Endlagenschalter detektieren die Endlage der X-Achse und vier Endlagenschalter die Endlage der Y-Achse.

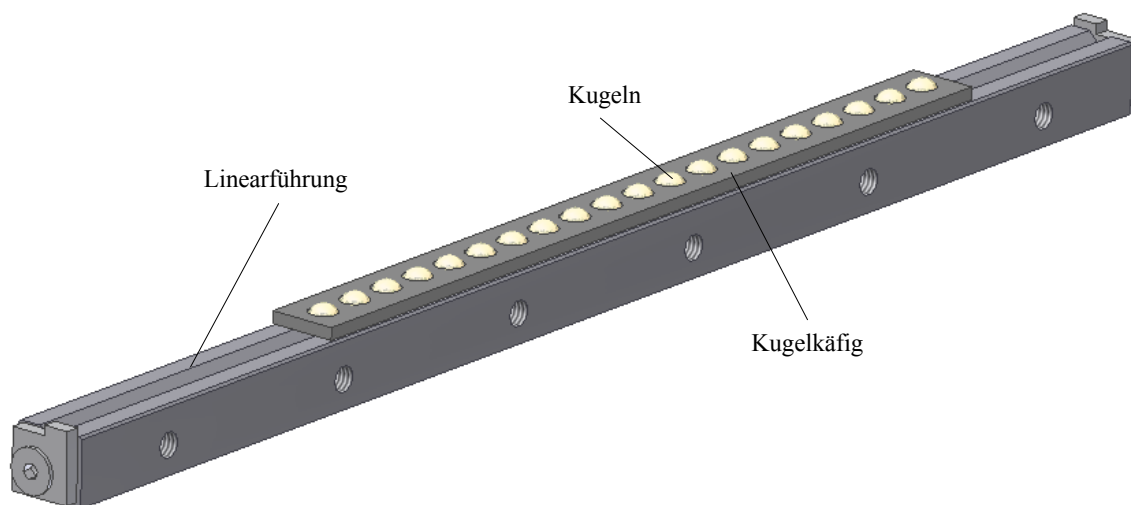


Abbildung 2.4: Verwendete Präzisionskugelführungen der Schneeberger AG

2.2 Sensorsystem

Zur Längenmessung werden zwei Planspiegelinterferometer der Firma SIOS Messtechnik GmbH eingesetzt. Diese homodynen Michelsoninterferometer wurden in Kooperation mit dem Institut für Prozessmess- und Sensortechnik der Technischen Universität Ilmenau entwickelt und verfügen über eine Auflösung von 0,1 nm [62, 65, 115]. Die verwendeten Miniaturinterferometer des Typs SP-2000 bestehen aus einem Messkopf sowie einer Versorgungs- und Auswerteeinheit (siehe Abbildung 2.5).



Abbildung 2.5: Planspiegelinterferometer der Serie SP bestehend aus a) Messkopf und b) Versorgungs- und Auswerteeinheit

Der Messkopf wird durch einen Lichtwellenleiter mit dem Licht eines frequenzstabilisierten HeNe Lasers gespeist, der sich in der Versorgungs- und Auswerteeinheit befindet und eine Wellenlänge von 632,9914 nm sowie eine Frequenzstabilität von $2 \cdot 10^{-8}$ besitzt [113, 117]. Diese räumliche Trennung von Laserrohr und Messkopf ist in der Präzisionslängenmesstechnik absolut notwendig, denn Laserrohre stellen Wärmequellen dar, die eine meist negative Auswirkung auf die Messbedingungen haben [110]. Der prinzipielle Aufbau des Messkopfes ist in Abbildung 2.6 dargestellt. Das über einen Lichtwellenleiter eingekoppelte Licht des stabilisierten Lasers wird durch einen polarisationsoptischen Strahlteiler in einen Mess- und einen Referenzstrahl geteilt. Nach der Reflexion am Mess- bzw. Referenzspiegel passieren beide Teilstrahlen erneut den Strahlteiler und interferieren nach einer $\lambda/4$ -Platte beim Austritt aus dem Teilerwürfel. Nach einer Blende passieren die Teilstrahlen eine Anordnung aus optischen und elektronischen Bauelementen, in welcher vier um jeweils 90 Grad phasenverschobene sinusförmige Messsignale von vier Fotoempfängern detektiert werden. Diese vier analogen Signale repräsentieren das Interferenzbild der interferierenden Teilstrahlen und stellen somit das Längenmesssignal dar.

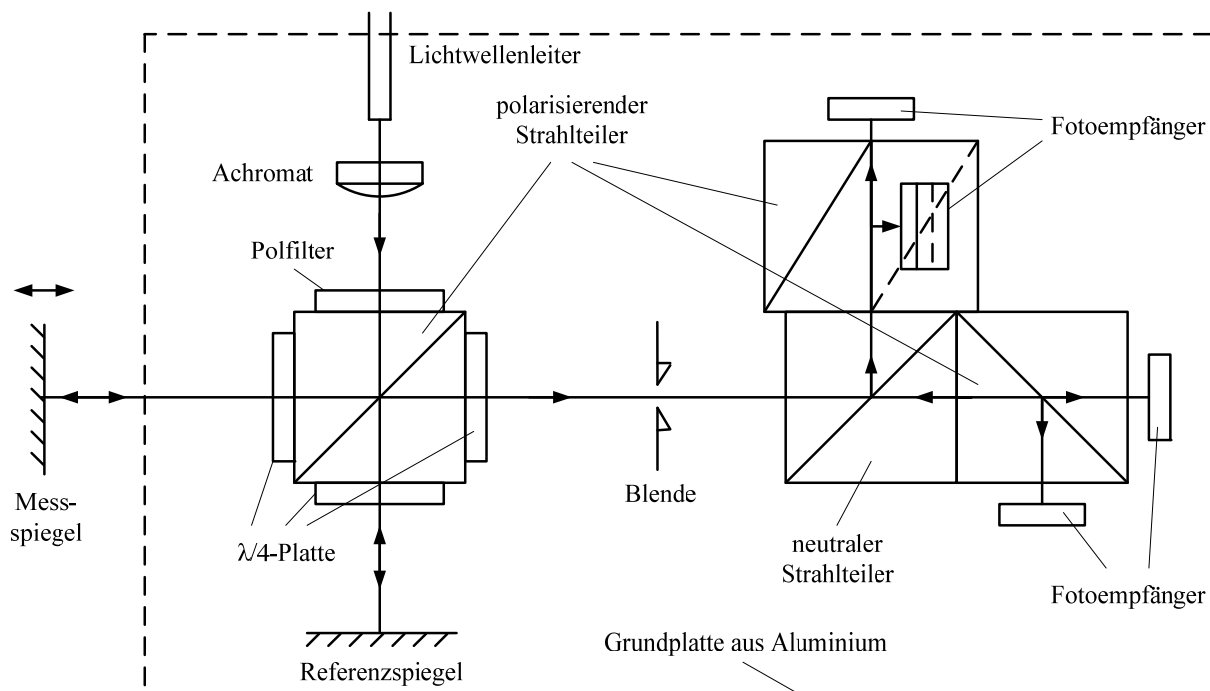


Abbildung 2.6: Prinzipieller Aufbau eines Planspiegelinterferometers (Draufsicht auf den Messkopf)

Nach einer Vorverstärkung im Messkopf werden die analogen Signale zur Versorgungs- und Auswerteeinheit geleitet. Durch Differenzbildung der komplementären Signale entstehen die Quadratursignale U_A und U_B , welche durch ein Reglermodul ausgewertet werden. Hier erfolgt eine nochmalige Verstärkung der Quadratursignale und eine Regelung von Amplitude und Offset [44]. Werden diese Signale anschließend auf einem Oszilloskop gegeneinander dargestellt, ergibt sich die für Interferometer typische kreisförmige Lissajous-Figur. Dabei entspricht ein Winkel von 360 Grad einer Bewegung des Messspiegels um die Hälfte der verwendeten Laserwellenlänge. Zusätzlich enthält die Versorgungs- und Auswerteeinheit eine Tachocontrollerkarte vom Typ DP-01, welche ebenfalls durch die Firma SIOS Messtechnik GmbH zur Verfügung gestellt wurde. Diese Karte wandelt die analogen Sinus- und Kosinus-signale in ein demoduliertes digitales Signal um, indem die analogen Signale gewandelt und die durchlaufenden Inkremente gezählt werden. Neben den analogen Signaleingängen verfügt die Karte über zwei digitale Eingänge zur Steuerung. Dies ist zum einen ein CLOCK-Signal zur Triggerung der Karte und zum anderen ein ZERO-Signal zum Rücksetzen des internen Zählers. Mit der steigenden Flanke des CLOCK-Signals wird der aktuelle Zählerstand in ein spezielles Register geschrieben und kann dort nach 200 ns von der nachgeschalteten Hardware übernommen werden. Bei dem Ausgangssignal handelt es sich um einen 32-Bit Integer Wert, der in einen Längenmesswert umgerechnet werden muss. Da die Quadratursignale mit einer Auflösung von 12-Bit gewandelt werden und somit die Position des Drehzeigers auf der Lissajous-Figur in 4096 Inkremente unterteilt wird, kann der Längenmesswert wie folgt berechnet werden:

$$l = N \cdot \frac{\lambda}{2} \cdot \frac{1}{4096} \quad (2.1)$$

Dabei ist N der ausgelesene Zählerstand der Tachocontrollerkarte und λ die Laserwellenlänge in Luft. Die berechnete Länge l ist die Differenzlänge zur Initialposition, denn Interferometer können prinzipbedingt nur relative Längenänderungen messen.

2.3 Antriebssystem

Die Achsen des in dieser Arbeit betrachteten NPM-Systems werden von eisenlosen Linear-motoren angetrieben. Diese Motoren basieren auf der Kraftwirkung, die auf bewegte Ladungen im magnetischen Feld ausgeübt wird [45]. Besonders in Anwendungen, in denen sowohl hohe Dynamik als auch extreme Präzision gefordert wird, bietet sich der Einsatz solcher Antriebe an. Dies liegt darin begründet, dass die Kraft in Richtung der Bewegung erzeugt wird und somit ein Bewegungsumformer nicht notwendig ist. Demzufolge wird das Verhalten des Antriebssystems nicht negativ durch Reibung, Spiel oder Elastizitäten beeinflusst und auch die zu bewegenden Massen können minimiert werden. Aufgrund dieser konzeptbedingten Vorteile resultiert bei elektrodynamischen Direktantrieben ein nahezu lineares Systemverhalten, denn die Kraft, die vom Motor aufgebracht wird, ist proportional zum Strom, der durch die Spule fließt [118]. Auch Grundlagenuntersuchungen hinsichtlich eines geeigneten Antriebssystems für NPM-Systeme mit großen Bewegungsbereichen konnten diese Einschätzung in der praktischen Anwendung untermauern [49]. Im Gegensatz zu [49] werden in dem hier

betrachteten NPM-System jedoch eisenlose Linearantriebe in Heteropolarausführung eingesetzt, denn aufgrund des sehr großen Bewegungsbereiches von 200 mm können Motoren in Homopolarausführung nicht verwendet werden. Der prinzipielle Aufbau der eingesetzten Motoren ist schematisch in Abbildung 2.7 dargestellt. Die felderzeugenden Permanentmagnete und zugehörigen Flussführungen im feststehenden Stator werden als Sekundärteil und der Läufer mit den stromdurchflossenen Spulen als Primärteil bezeichnet [118]. Der Sekundärteil besteht aus einem U-förmigen Weicheisenprofil, das mit abwechselnd gepolten Permanentmagneten bestückt ist. Diese Anordnung bewirkt die Ausbildung eines sinusförmigen Dauermagnetfeldes entlang der Bewegungsachse. Der Primärteil besteht aus zwei in Epoxidharz eingegossenen Kurzspulen und wird berührungslos im Sekundärteil platziert. Abbildung 2.8 zeigt ein CAD-Modell der eingesetzten Linearantriebe vom Typ ULIM3-2P-66, welche von der Firma IDAM hergestellt werden [56]. Es ist deutlich zu erkennen, dass der Stator mindestens so lang wie der Bewegungsbereich sein muss. Somit ergibt sich ein vergleichsweise großer Motor. Aufgrund des sinusförmigen Dauermagnetfeldes im Stator verlässt jede Teilspule beziehungsweise jeder Spulenstrang bei einer größeren Bewegung die Luftspalte einer Feldrichtung und fährt in entgegengesetzt gepolte Feldgebiete ein. Dieser Umstand erfordert die Umkehr der Stromrichtung (Kommutierung) in jedem Strang, denn nur so kann eine konstante Kraft in Bewegungsrichtung erzeugt werden. Um eine elektronische Kommutierung zu realisieren, befindet sich in jeder Kurzspule ein integrierter Hallsensor, welcher die magnetische Flussdichte B misst. Auf Basis dieser Messgrößen reguliert das Echtzeitsystem die Spulenströme derart, dass in Summe die gewünschte Antriebskraft erzeugt wird. Die zwei Spulen sind um ein Viertel der Wellenlänge des Permanentmagnetfeldes verschoben angeordnet. Damit ist gewährleistet, dass sich immer eine der beiden Spulen im Krafteingriff befindet.

Neben dem eigentlichen Linearmotor besteht das Antriebssystem aus zwei analogen Leistungsverstärkern pro Motor, welche die beiden Spulen des Primärteiles bestromen. Vom Steuerungs- und Regelungssystem werden Steuersignale in Form von Spannungen zwischen ± 10 V vorgegeben, die von den Leistungsendstufen in einen proportionalen Strom von $\pm 1,5$ A gewandelt werden.

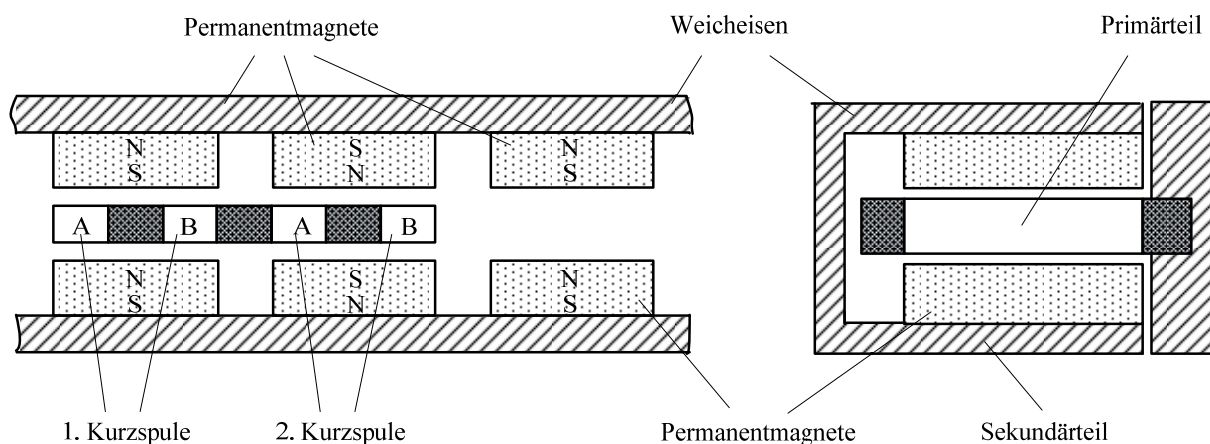


Abbildung 2.7: Prinzipieller Aufbau der verwendeten Linearantriebe (Schnittdarstellung)

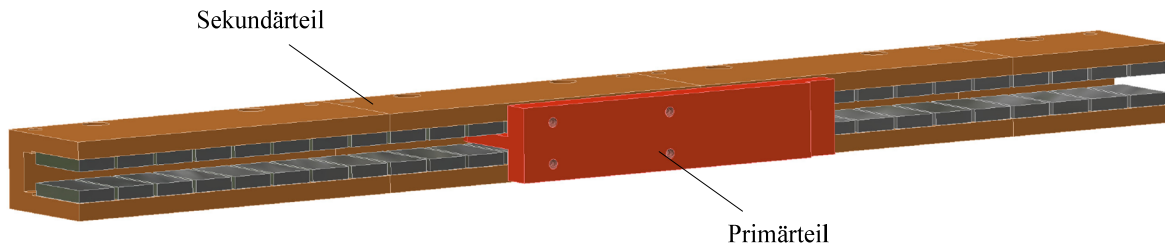


Abbildung 2.8: CAD-Modell der verwendeten Linearantriebe vom Typ ULIM3-2P-66

Um Abweichungen des Sollstromes, ausgelöst durch Selbstinduktion in den Spulen, zu kompensieren, verfügen die Verstärker über integrierte Stromregler. Des Weiteren besitzt ein Endstufenverstärker zwei digitale Eingänge, die zur Freischaltung des Verstärkers sowie zur Signalisierung von Fehlfunktionen dienen. Wie Abbildung 2.9 a) zeigt, sind die Motorenverstärker als Leiterkarten in einem Format von 160 mm x 100 mm ausgeführt. Jeweils vier dieser Karten bilden eine Motoreinheit zum Betrieb einer Bewegungsachse (siehe Abbildung 2.9 b)).

Wie eingangs bereits angedeutet, arbeiten die verwendeten eisenlosen Linearmotoren aufgrund des berührungsfreien Betriebes ohne mechanischen Verschleiß. Die Antriebe sind auf hohe dynamische Anforderungen ausgelegt und können laut Datenblatt Geschwindigkeiten von mehr als 2 m/s und Beschleunigungen von über 20 m/s^2 erreichen [56].



Abbildung 2.9: a) Leistungsverstärkerkarte sowie b) Verstärkereinheit für eine Achse in 19'' Ausführung

2.4 Positionsinitialisierung

Nachteilig an dem in Abschnitt 2.2 vorgestellten interferenzoptischen Messprinzip ist dessen Eigenschaft, dass lediglich relative Positionsänderungen detektiert werden können. Somit wird ein Referenzsystem benötigt, mit dem auf die absolute Position geschlossen werden kann. Das Grundprinzip des hier eingesetzten Referenzsystems kann mit einer Art Fadenkreuz verglichen werden, was es ermöglicht, einen zweidimensionalen Initialisierungspunkt bis auf wenige Nanometer genau zu reproduzieren [68]. Der Aufbau des Referenzsystems besteht aus einem Referenzgestell und einer so genannten Fahne. In dem Referenzgestell sind zwei um 90 Grad versetzte Zweiquadrantenphotoempfänger und zwei Dioden installiert. Abbildung 2.10 zeigt ein CAD-Modell des Referenzgestelles sowie der Fahne. In der Fahne sind zwei Spalte mit einer Breite von 0,5 mm eingebracht, die orthogonal zueinander stehen. Der horizontale Spalt dient dem Referenzsystem zur Positionierung der X-Achse und der vertikale Spalt wird für die Positionierung der Y-Achse benötigt. Die Fahne ist am Ende des X-Sliders angebracht

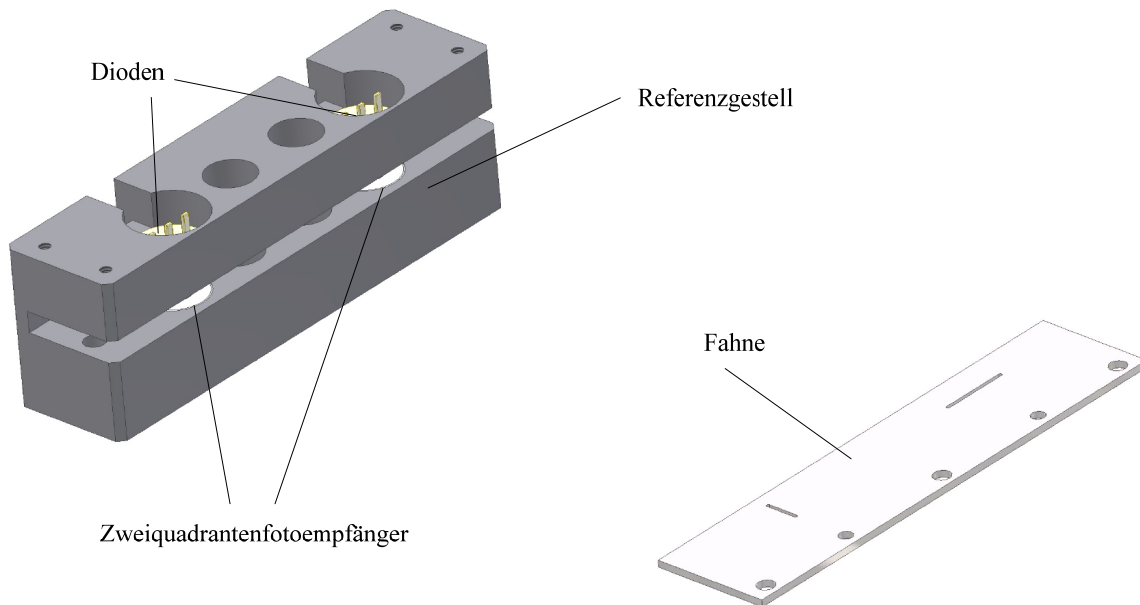


Abbildung 2.10: Komponenten des Positionsreferenzsystems (CAD-Modell)

und besitzt somit einen definierten Bezug zur Spiegelecke. Befindet sich die Fahne an einer bestimmten Position im Referenzgestell, fällt durch die zwei Spalte das Licht der Dioden. Die Intensität des Lichtes wird auf der gegenüberliegenden Seite durch die Zweiquadrantenphotoempfänger detektiert und anhand der Intensitätsverteilung des Lichtes kann auf die absolute Position der Fahne geschlossen werden. Grund für den Einsatz von Zweiquadrantenfotoempfängern ist der Umstand, dass das Differenzsignal zwischen den beiden Quadranten zur Auswertung genutzt werden kann. Somit sind Störeinflüsse, die auf beide Quadranten wirken, wie beispielsweise Alterungsprozesse, produktionsbedingte Qualitätsschwankungen oder Umwelteinflüsse ausgeschlossen.

Beim Passieren eines Spaltes entsteht der in Abbildung 2.11 dargestellte charakteristische Signalverlauf. In dieser Grafik ist zu erkennen, dass das Differenzsignal nur in einem Bereich von circa 4 mm von Null verschieden ist.

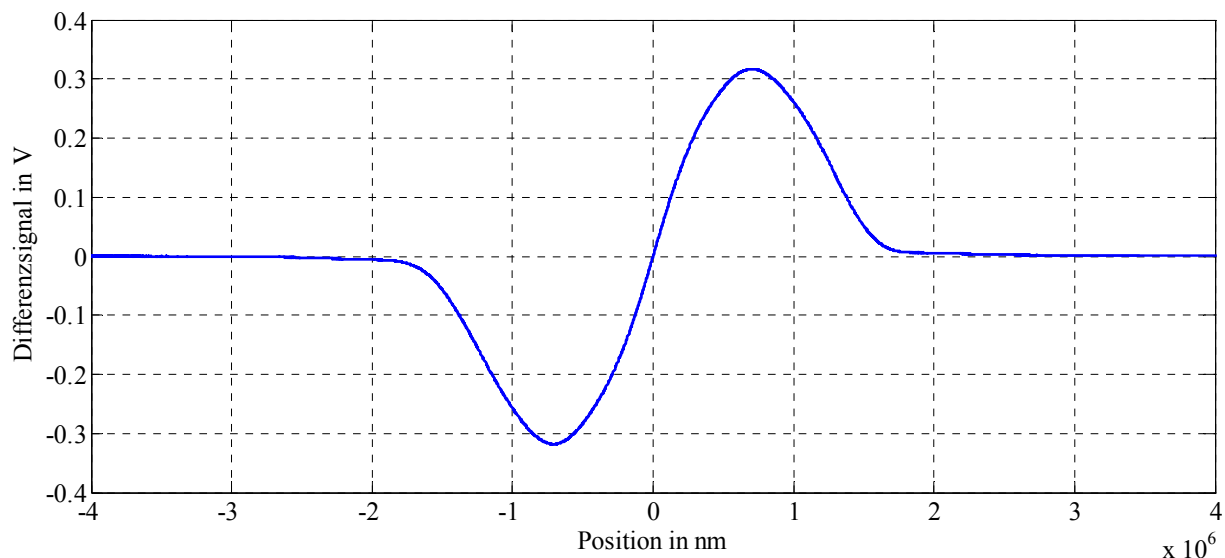


Abbildung 2.11: Differenzsignal der Zweiquadrantenphotoempfänger beim Passieren des Spaltes

In dieser Region treten zwei Extrema auf, die sich durch Abschattung eines der beiden Quadranten ergeben. Null ist das Differenzsignal genau dann, wenn sich die Fahne nicht im Referenzgestell befindet, beide Quadranten gleichmäßig von der Fahne verdeckt werden oder der Spalt genau unter der Diode steht und das Licht mit maximaler Intensität beide Quadranten bestrahlt. Der letztgenannte Fall tritt nur in einem Bereich von $> 200 \text{ nm}$ auf und somit kann dieser Punkt als reproduzierbare Referenz genutzt werden [68]. Das von den Fotoempfängern erzeugte Differenzsignal wird auf $\pm 10 \text{ V}$ verstärkt und in das Echtzeitsystem (siehe Abschnitt 2.5) eingelesen. Auf Basis dieser Messwerte wurde im Rahmen von [68] ein Regelungskonzept erarbeitet, welches es erlaubt, den Initialisierungspunkt automatisiert anzufahren und auf wenige Nanometer genau zu reproduzieren.

2.5 Echtzeithardware

Die Regelung des experimentellen Aufbaus wird durch ein modulares Rapid Control Prototyping System der Firma dSPACE realisiert [38]. Dieses besteht aus einer Hardware- und einer Softwarekomponente, die nicht unabhängig voneinander einsetzbar sind. Wie Abbildung 2.12 zeigt, besteht der Hardwareteil des Systems aus einer DS1006 Prozessorkarte sowie verschiedenen Erweiterungskarten für die digitale sowie analoge Ein- und Ausgabe von Signalen. Alle Hardwarekomponenten sind in einer dSPACE Erweiterungsbox des Typs PX20 untergebracht.

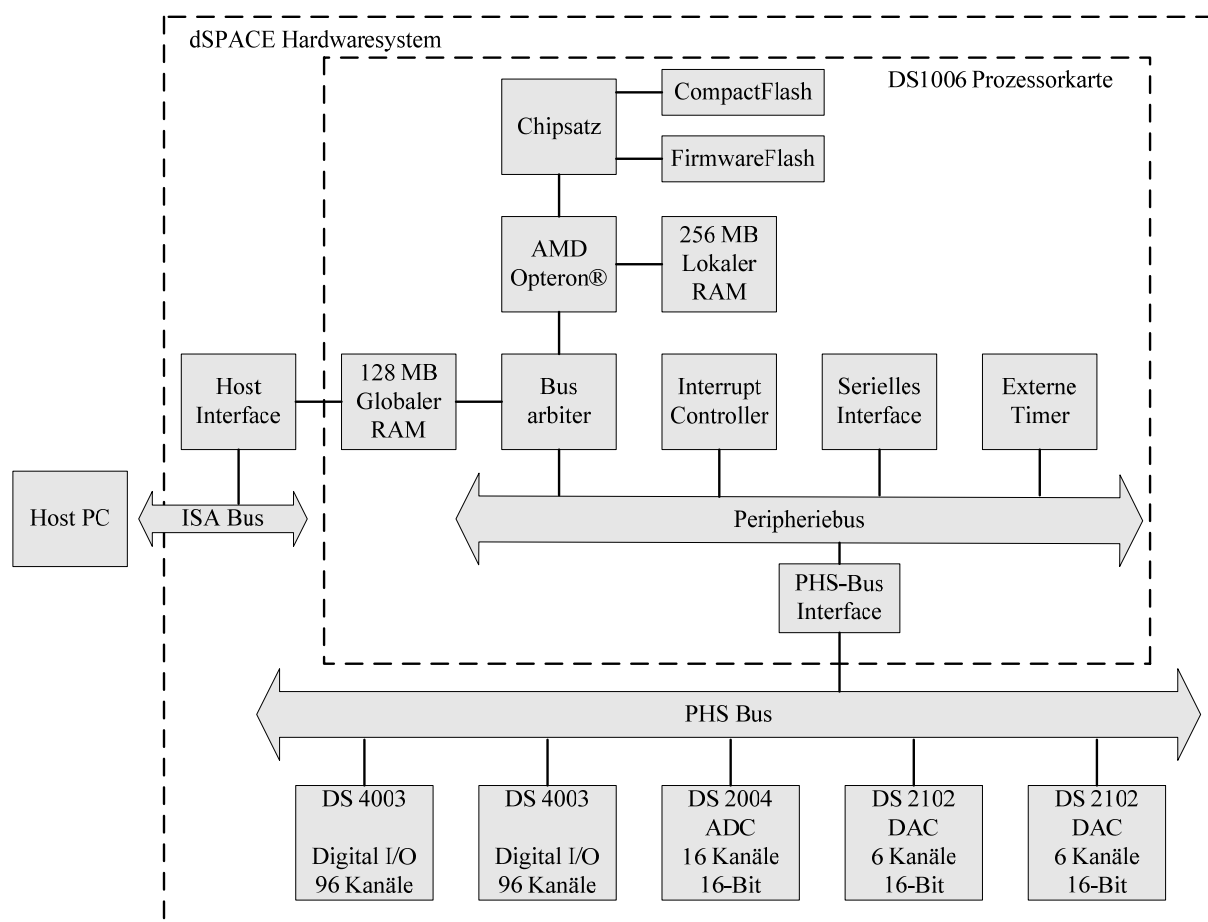


Abbildung 2.12: Hardwarestruktur des modularen Rapid Control Prototyping Systems

Als Prozessor wird ein AMD Opteron® Prozessor mit einer Taktrate von 3,0 GHz eingesetzt. Die Prozessorkarte steuert die angeschlossenen I/O Karten über einen so genannten PHS Bus und berechnet die Regelalgorithmen. Wie bereits in Abschnitt 2.2 beschrieben, werden die Positionswerte aus den Registern der Tachocontrollerkarten über zwei digitale Schnittstellenkarten des Typs DS4003 am Echtzeitsystem eingelesen. Zur gleichen Zeit werden auch die acht Hallsensorsignale und die zwei Differenzsignale des Referenzsystems über eine A/D Wandlerkarte vom Typ DS2004 mit einer 16-Bit Auflösung aufgenommen. Anschließend werden die Positionswerte, die Hallsensorsignale sowie die Differenzsignale über das proprietäre PHS Bussystem an die Prozessorkarte übertragen und der Regelalgorithmus berechnet die für den aktuellen Tastschritt gültigen Stelleingriffe. Die Stellsignale werden an zwei D/A Wandlerkarten des Typs DS2102 transferiert und mit einer 16-Bit Auflösung an die Leistungsverstärker ausgegeben. Der beschriebene Vorgang wird in einem Zyklus von 100 μ s ausgeführt, was einer Samplingfrequenz von 10 kHz entspricht. Neben dem beschriebenen zyklischen Regelalgorithmus werden zwei weitere ereignisdiskrete Prozesse durch das Echtzeitsystem gesteuert. Zum einen ist dies die benutzergesteuerte Freischaltung der Leistungsverstärker. Angesteuert werden die Motoreinheiten durch jeweils vier digitale Signale, die es ermöglichen, jede Leistungsstufe separat freizuschalten. Zum anderen werden in jedem Tastschritt alle sechs Endlagenschalter ausgewertet. Auch dies erfolgt mithilfe sechs digitaler Eingänge auf einer der DS4003 Karten. Zur Verdeutlichung sind in Abbildung 2.13 die beschriebenen Signalwege nochmals schematisch dargestellt.

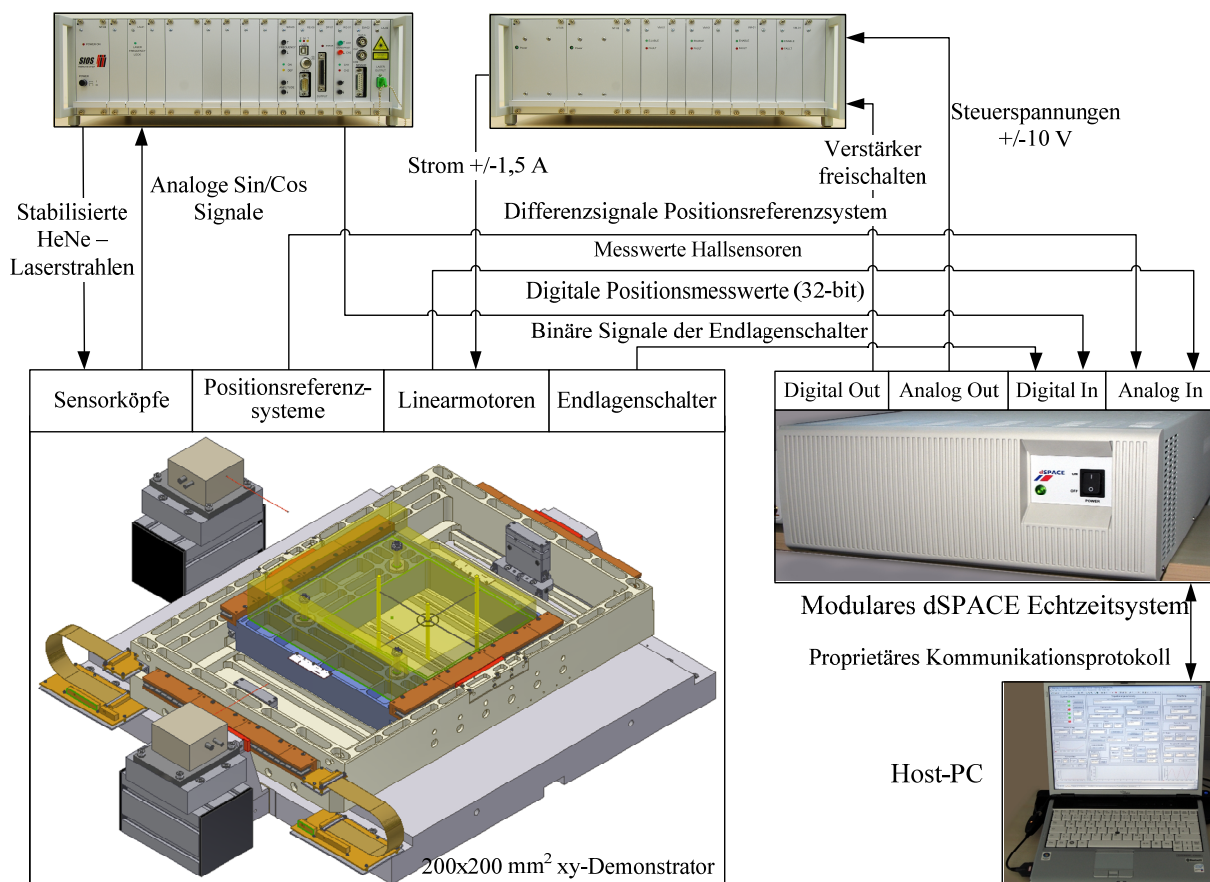


Abbildung 2.13: Signalfluss zwischen dem Echtzeitsystem und dem zweiachsigen Demonstrator

Der eigentliche Regelalgorithmus wird auf einem Host-PC in Matlab/Simulink® erstellt [123]. Hierzu stehen in Simulink® umfangreiche Bibliotheken in Form von parametrierbaren Blöcken zur Verfügung, mit denen das Regelungssystem modelliert werden kann. Um den Zugriff auf die Hardware des Rapid Control Prototyping Systems zu ermöglichen, wird über das Real-Time-Interface von Simulink® durch dSPACE eine spezielle Bibliothek zur Verfügung gestellt. Nach der Entwicklung der Regelalgorithmen wird das Simulink®-Modell automatisiert mithilfe des integrierten Real-TimeWorkshop [124] in ANSI C Code übersetzt und dieser wird im Anschluss an einen Crosscompiler von dSPACE übergeben. Ist die Kompilierung erfolgreich abgeschlossen, wird der Maschinencode über eine PCMCIA Schnittstelle auf das Echtzeitsystem geladen. Ab diesem Zeitpunkt wird der Regelalgorithmus in Echtzeit ausgeführt. Mit der Software ControlDesk von dSPACE lassen sich im laufenden Betrieb die Parameter des Regelalgorithmus vom Host-PC aus verändern. Zusätzlich ist eine Visualisierung und Speicherung aller im Regelungssystem auftretender Signale möglich. Somit bietet das dSPACE System die Möglichkeit, schnell und unkompliziert neue Regelungsalgorithmen am realen mechatronischen System zu erproben.

3 Modellbildung

Gegenstand dieses Kapitels ist die mathematische Beschreibung des dynamischen Verhaltens des in Kapitel 2 vorgestellten Nanopositioniersystems. Auch der Modellbildungsprozess beginnt mit dem mechanischen Teil des experimentellen Aufbaues, wobei die stark nichtlineare Reibkraft vorerst vernachlässigt wird. In den darauffolgenden Abschnitten 3.2 und 3.3 findet die Modellierung des Sensor- sowie des Aktorsystems statt. Der Schwerpunkt des Kapitels liegt auf der modellbasierten Beschreibung der Reibkraft. Deren Einfluss auf das Systemverhalten sowie deren Verhalten im Nanometerbereich wird ausführlich in Abschnitt 3.4 abgehandelt.

3.1 Dynamisches Verhalten des mechanischen Aufbaus

Da das in dieser Arbeit betrachtete Nanopositioniersystem konstruktionsbedingt nur minimale Verkopplungen zwischen X- und Y-Achse aufweist², besteht die Möglichkeit, das dynamische Verhalten der Bewegungsachsen unabhängig voneinander zu modellieren. Weiterhin kann eine Achse des Systems als Punktmasse angesehen werden, welche durch die Antriebskraft $F_A(t)$ beschleunigt und von der Reibkraft $F_R(t)$ abgebremst wird. Daher kann unter Nutzung des zweiten Newton'schen Axioms das dynamische Verhalten wie folgt beschrieben werden [45]:

$$m \cdot \ddot{x}(t) + F_R(t) = F_A(t) \quad (3.1)$$

Aufgrund des dominanten Einflusses der Reibung auf das Verhalten von NPM-Systemen wird diese unabhängig von der Trägheit des Systems in Abschnitt 3.4 modelliert. Zunächst gilt unter der Annahme, dass keine Reibung im System wirkt:

$$m \cdot \ddot{x}(t) = F_A(t) \quad (3.2)$$

Werden die Beschleunigung $\ddot{x}(t)$ und die Geschwindigkeit $\dot{x}(t)$ als die inneren Zustände des Systems angesehen, kann Gleichung (3.2) in den Zustandsraum überführt werden [43]:

$$\begin{bmatrix} \dot{x}(t) \\ \ddot{x}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{1}{m} \end{bmatrix} \cdot F_A(t) \quad (3.3)$$

² An dieser Stelle soll eine leichte Verkopplung der Bewegungsachsen angenommen werden. Ein experimenteller Beweis der These erfolgt in Kapitel 6.

$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix}$$

Mit Gleichung (3.3) liegt nun ein dynamisches Modell des mechanischen Teils einer Bewegungsachse vor. Dieses wird im weiteren Verlauf der Arbeit sowohl für die X- als auch für die Y-Achse Anwendung finden. Die Masse m der X-Achse beträgt hierbei 13 kg und die Masse der Y-Achse 24,7 kg.

3.2 Dynamisches Verhalten des Sensorsystems

Die Position des betrachteten Nanopositioniersystems wird mit Interferometern erfasst und die Längenmesswerte werden bereits von den Versorgungs- und Auswerteeinheiten in Form digitaler 32 Bit Werte zur Verfügung gestellt. Wie schon in Abschnitt 2.2 dargestellt, kann alle 200 ns ein neuer Messwert vom Echtzeitsystem ausgelesen werden. Dies entspricht einer maximalen Samplingfrequenz von 500 kHz, und somit ist die Dynamik des Sensorsystems um Größenordnungen höher als die Dynamik der verbleibenden Systemkomponenten. Auf Grund dieser Tatsache kann eine Modellbildung der Sensordynamik an dieser Stelle unterbleiben, da ein nahezu verzögerungsfreies Übertragungsverhalten angenommen werden kann.

3.3 Dynamisches Verhalten des Aktorsystems

Wie bereits in Abschnitt 2.3 erläutert wurde, setzt sich das Antriebssystem aus einem Linear-motor und den dazugehörigen Leistungsendstufen zusammen. Die von den Linearantrieben erzeugte Antriebskraft ist identisch mit der Kraft (Lorentzkraft), die auf bewegte Ladungen im Magnetfeld wirkt. Zur Verdeutlichung ist der beschriebene Zusammenhang in Abbildung 3.1 dargestellt. Aus der elektrodynamischen Kraftwirkung auf bewegte Ladungen folgt, dass im quasistationären Magnetfeld der Flussdichte B auf einen Leiter, der vom Strom $i(t)$ durchflossen wird und eine kraftwirksame Leiterlänge l besitzt, eine Kraft $F_E(t)$ wirkt [118].

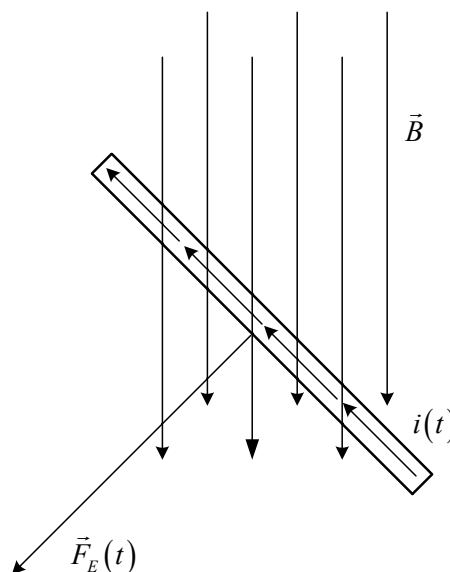


Abbildung 3.1: Kraft auf einen stromdurchflossenen Leiter im Magnetfeld

Stehen die drei Größen senkrecht aufeinander, gilt:

$$F_E(t) = B \cdot l \cdot i(t) = k_{LM} \cdot i(t) \quad (3.4)$$

Bei Linearantrieben wird k_{LM} als Kraft- oder Motorkonstante bezeichnet. Wie in Abschnitt 2.3 beschrieben wurde, treiben Linearmotoren in Heteropolarausführung die Bewegungsachsen des Demonstrators an. Somit setzt sich der resultierende Kraftvektor aus den Teilkräften zusammen, die von den zwei Kurzspulen erzeugt werden. Dieser Zusammenhang kann allgemein wie folgt formuliert werden:

$$\begin{aligned} F_{LM}(t) &= F_{E1}(t) + F_{E2}(t) \\ &= i_1(t) \cdot B_1 \cdot l_1 + i_2(t) \cdot B_2 \cdot l_2 \end{aligned} \quad (3.5)$$

Unter der Voraussetzung, dass die integrierten Hallsensoren die wirksame Flussdichte exakt bestimmen und dadurch eine präzise Kommutierung der Spulenströme möglich ist, gilt für die eingesetzten Motoren:

$$F_{LM}(t) = k_{LM} \cdot (i_1(t) + i_2(t)) = k_{LM} \cdot i_A(t) \quad (3.6)$$

Die Kraftkonstante k_{LM} ist hierbei laut Herstellerangaben 6,2 N/A [56]. Anhand von Gleichung (3.6) ist klar zu erkennen, dass über eine Regulierung des Spulenstromes $i_A(t)$ eine stromproportionale Antriebskraft erzeugt werden kann. Da die Spulen im Primärteil des Motors durch die zugehörigen Leistungsverstärker bestromt werden, muss im zweiten Schritt der Systemidentifikation die Dynamik der Endstufen untersucht werden. Hierzu wird eine Verstärkerkarte mit sinusförmigen Spannungssignalen angeregt und der zur Ansteuerung des Linearmotors ausgegebene Strom messtechnisch erfasst. Als Last werden Spulen eingesetzt, deren elektrisches Verhalten identisch ist mit den im Primärteil vergossenen Kurzspulen. Abbildung 3.2 zeigt das dynamische Verhalten der Leistungsverstärker. Grund für die konstante Verstärkung von -16,47 dB ist die Tatsache, dass die Leistungsstufen Ströme in einem Bereich von $\pm 1,5$ A ausgeben und über einen hochohmigen Eingang vom Echtzeitsystem (siehe Abschnitt 2.5) in einem Bereich von ± 10 V angesteuert werden.

Somit ergibt sich ein konstantes Verstärkungsverhältnis der Leistungsverstärker von:

$$k_V = \frac{i_{max}}{u_{max}} = 0,15 \frac{A}{V} \triangleq -16,47 \text{ dB} \quad (3.7)$$

Weiterhin ist zu sehen, dass die Verstärkung der Karte bis zu einer Frequenz von etwa 10 kHz nahezu konstant bleibt und anschließend eine sehr starke Dämpfung einsetzt. Weitere Untersuchungen zeigten, dass das beobachtete Verhalten mit einem PT₄-Glieder approximiert werden kann:

$$G(s) = \frac{k_V}{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + 1} \quad (3.8)$$

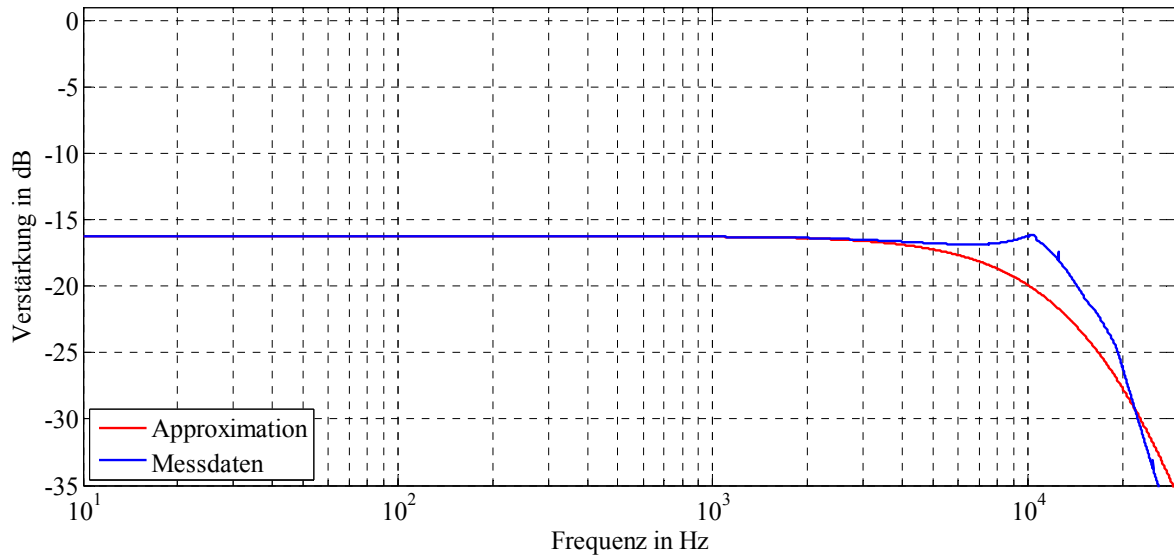


Abbildung 3.2: Amplitudengang der Leistungsverstärkerkarten

Die Parameter a_i in Gleichung (3.8) können durch eine Regression ermittelt werden und ergeben sich zu:

$$\begin{aligned} a_1 &= 3,072 \cdot 10^{-5} \quad s \\ a_2 &= 3,533 \cdot 10^{-10} \quad s^2 \\ a_3 &= 1,803 \cdot 10^{-15} \quad s^3 \\ a_4 &= 3,443 \cdot 10^{-21} \quad s^4 \end{aligned}$$

Da die Tastrate des Regelungssystems mit 10 kHz deutlich unterhalb der Grenzfrequenz der Verstärkerkarte liegt, kann aus regelungstechnischer Sicht die Übergangsfunktion der Endstufe als ein Proportionalelement angenommen werden. Auch kann davon ausgegangen werden, dass der vom Leistungsverstärker vorgegebene Strom sofort fließt und durch den integrierten hochdynamischen Stromregler konstant bleibt. Dies soll auch für die Selbstinduktion und in gewissem Maße auch für die Bewegungsinduktion gelten. Daher gilt für die Leistungselektronik eines Linearantriebes:

$$i_A(t) = k_V \cdot u_S(t) \quad (3.9)$$

Der Parameter $u_S(t)$ stellt in Gleichung (3.9) die vom Echtzeitsystem bereitgestellte Steuerungsspannung dar. Gemäß Gleichung (3.6) wird die Krafterzeugung der Linearmotoren durch die Spulenströme bestimmt und so kann das Übertragungsverhalten eines Linearmotors mit zugehöriger Leistungselektronik wie folgt approximiert werden:

$$F_{LM}(t) = k_V \cdot k_{LM} \cdot u_S(t) \quad (3.10)$$

Weiterhin entspricht die auf eine Bewegungsachse des Demonstrators wirkende Antriebskraft $F_A(t)$ dem Doppelten von $F_{LM}(t)$, da zwei Linearmotoren je Achse verbaut sind (siehe Abbildung 2.2). Demzufolge kann das Verhalten des Antriebssystems einer Bewegungsachse beschrieben werden durch:

$$F_A(t) = 2 \cdot k_{LM} \cdot i_A(t) = k_A \cdot k_V \cdot u_S(t) \quad \text{mit } k_A = 2 \cdot k_{LM} \quad (3.11)$$

Wird $F_A(t)$ in Gleichung (3.3) mit Gleichung (3.11) ersetzt, ergibt sich ein direkter Zusammenhang zwischen den Stellsignalen des Echtzeitsystems und der daraus resultierenden Position des Sliders. Somit kann unter Vernachlässigung der Reibung das dynamische Verhalten einer Bewegungsachse beschrieben werden durch:

$$\begin{bmatrix} \dot{x}(t) \\ \ddot{x}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_A \cdot k_V}{m} \end{bmatrix} \cdot u_s(t) \quad (3.12)$$

$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix}$$

3.4 Modellierung der Reibkraft

Wie bereits zu Beginn dieses Kapitels erwähnt wurde, dominiert die Reibung das Verhalten des Gesamtsystems. Um diese These experimentell zu belegen, wurde die Y-Achse mit einer Geschwindigkeit von 1 mm/s über eine Strecke von 10 mm verfahren³. Gemäß Gleichung (3.1) müssen die Antriebe des xy-Demonstrators bei diesem Experiment zum einen die träge Masse beschleunigen und zum anderen die Reibkräfte in den Kugelführungen überwinden. Somit kann die Reibkraft indirekt aus der Differenz zwischen Antriebs- und Beschleunigungskraft berechnet werden. Abbildung 3.3 zeigt für das beschriebene Experiment beide Größen. Anhand der abgebildeten Graphen ist klar zu erkennen, dass nur ein sehr geringer Teil der Antriebskraft für die Beschleunigung der trägen Masse aufgewendet wird. Über 95 %⁴ der Motorleistung dienen zur Überwindung der Reibkräfte, was deren dominanten Einfluss bei kugelgeführten NPM-Systemen unterstreicht. Somit ist es zwingend notwendig, im Zuge der Modellbildung den Schwerpunkt auf eine möglichst exakte Abbildung der Reibkraft zu legen.

Erschwerend kommt hinzu, dass die Reibung besonders im Nanometerbereich ein stark nicht-lineares Verhalten aufweist. Diese Aussage wird phänomenologisch im Abschnitt 3.4.1 überprüft und anhand von experimentell ermittelten Daten belegt. Im Anschluss daran wird in Abschnitt 3.4.2 der Wissensstand auf dem Gebiet einer regelungstechnisch angepassten Reibmodellierung ausführlich dargelegt. Basierend auf experimentellen Untersuchungen erfolgt in Abschnitt 3.4.3 die Parameteridentifikation von fünf ausgewählten dynamischen Reibmodellen. In diesem Abschnitt wird für jedes Modell der Parameteridentifikationsprozess vorgestellt und die erzielten Ergebnisse diskutiert. Abgeschlossen wird der Abschnitt mit einer Leistungsbewertung der identifizierten Modelle anhand von Validierungsdaten. Als Ergebnis des Kapitels werden zwei dynamische Reibmodelle für die spätere Verwendung in einem modellbasierten Regelungssystem ausgewählt (siehe Kapitel 4).

³ Hin- und Rückfahrt

⁴ Dieser Wert wurde aus dem Verhältnis der Maximalwerte bestimmt.

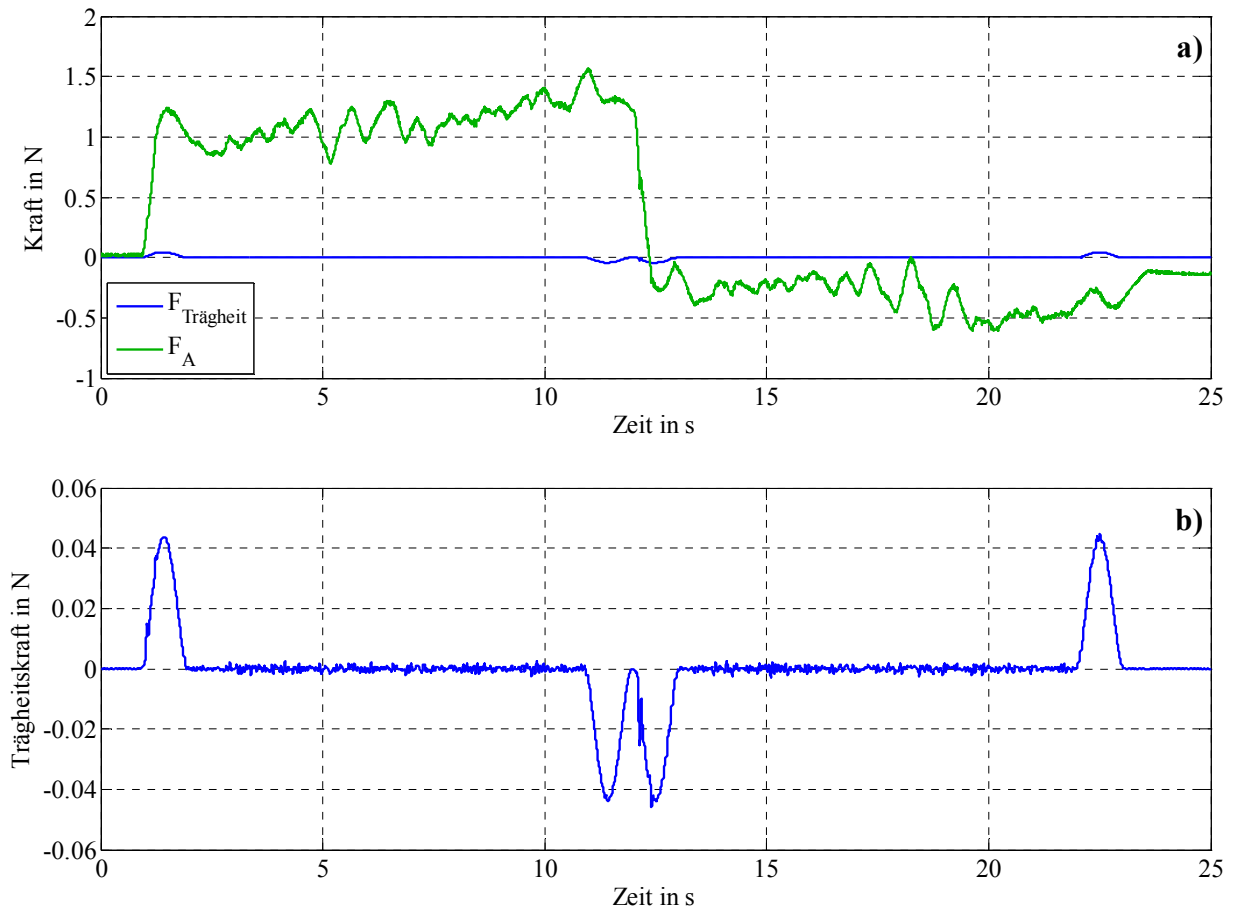


Abbildung 3.3: Einfluss der Reibung auf das Systemverhalten: (a) Gegenüberstellung von Antriebskraft und Trägheitskraft; (b) Vergrößerte Darstellung der Trägheitskraft

3.4.1 Reibphänomene

In nahezu allen mechanischen Systemen ist Reibung präsent. Diese wird verursacht durch die gegenseitige Interaktion der in Kontakt stehenden Oberflächen. Das resultierende Reibverhalten wird beeinflusst durch eine Vielzahl von Oberflächenparametern wie beispielsweise der Oberflächentopografie, den eingesetzten Oberflächenmaterialien oder dem verwendeten Schmiermittel. Prinzipiell lässt sich das Reibverhalten in zwei Domänen unterteilen, den so genannten „Pre-Sliding“- („micro-slip“) und den so genannten „Sliding“- („gross-sliding“) Bereich [12, 95, 135]. De Moerlooze und Al-Bender konnten in [2, 32] zeigen, dass diese Eigenschaften auch für Kugelführungen mit V-Nuten gelten. Somit wird im Folgenden anstelle von „Sliding“⁵ der englische Begriff „Rolling“ Anwendung finden.

Im „Pre-Rolling“-Bereich haften die interagierenden Oberflächen an den Kontaktstellen aufgrund von atomaren Wechselwirkungen aneinander. Wirkt eine begrenzte tangential Kraft, so werden die reibenden Körper gegeneinander verschoben. Die verrichtete Arbeit wird an den Kontaktflächen gespeichert und somit wächst die Rückstellkraft proportional mit der Auslenkung. Wird die angreifende Kraft wieder reduziert, begeben sich die beiden Körper in

⁵ Dieser Begriff wird in der Literatur im Zusammenhang mit ungelagerter Reibung gebraucht.

ihre Ausgangsposition. Mit der Zielstellung, dieses in der Literatur beschriebene Verhalten auch am xy-Demonstrator nachzuvollziehen, wird ein angepasstes Experiment durchgeführt. Die Y-Achse des Demonstrators wird mit einer Folge aus dreiecksförmigen Stromimpulsen beaufschlagt (siehe Abbildung A 1 in Anhang A) und die resultierende Positionsveränderung ausgewertet (siehe Abbildung A 2). Da im „Pre-Rolling“-Bereich aufgrund der geringen Beschleunigungen Trägheitskräfte vernachlässigt werden können, gilt nach Gleichung (3.1) $F_A(t) = F_R(t)$. In Abbildung 3.4 ist zur Verdeutlichung die Reibkraft in Abhängigkeit der Position aufgetragen. Es ist klar zu erkennen, dass das in der Literatur beschriebene Verhalten auch am xy-Demonstrator beobachtet werden kann. Die Bewegung ist zum einen vollständig reversibel und zum anderen ist der Betrag der Reibkraft abhängig von der Bewegungsrichtung. Darüber hinaus kann anhand der Abbildung festgestellt werden, dass das Reibverhalten ausgesprochen gut reproduzierbar ist, denn auf identische Anregungen folgen identische Systemreaktionen.

Abbildung 3.5 zeigt eine weitere Eigenschaft der Reibung im „Pre-Rolling“-Bereich. Wird das System im geregelten Betrieb unterschiedlich weit ausgelenkt (siehe Abbildung A 3), beschreibt die Reibkraft (siehe Abbildung A 4) immer den gleichen Pfad im Kraft-Positions-Diagramm. Diese Eigenschaft ist unter dem Begriff „hysteresis with non-local memory“ bekannt [32] und kann auch bei NPM-Systemen nachgewiesen werden.

Überschreitet die angreifende Kraft einen systemspezifischen Grenzwert, so beginnen die Oberflächen teilweise übereinander zu gleiten und es kommt zu Bewegungen mit nicht reversiblen (plastischen) Anteilen. Wird die tangentielle Kraft weiter gesteigert, reißen alle Verbindungen zwischen den Kontaktstellen und der elastische Anteil der Bewegung bleibt konstant. Nun kommt es zu rein plastischen Bewegungen und das System befindet sich im „Rolling“-Bereich. Auch das Verhalten im beschriebenen Übergangsbereich wird anhand ausgewählter Experimente am xy-Demonstrator nachvollzogen.

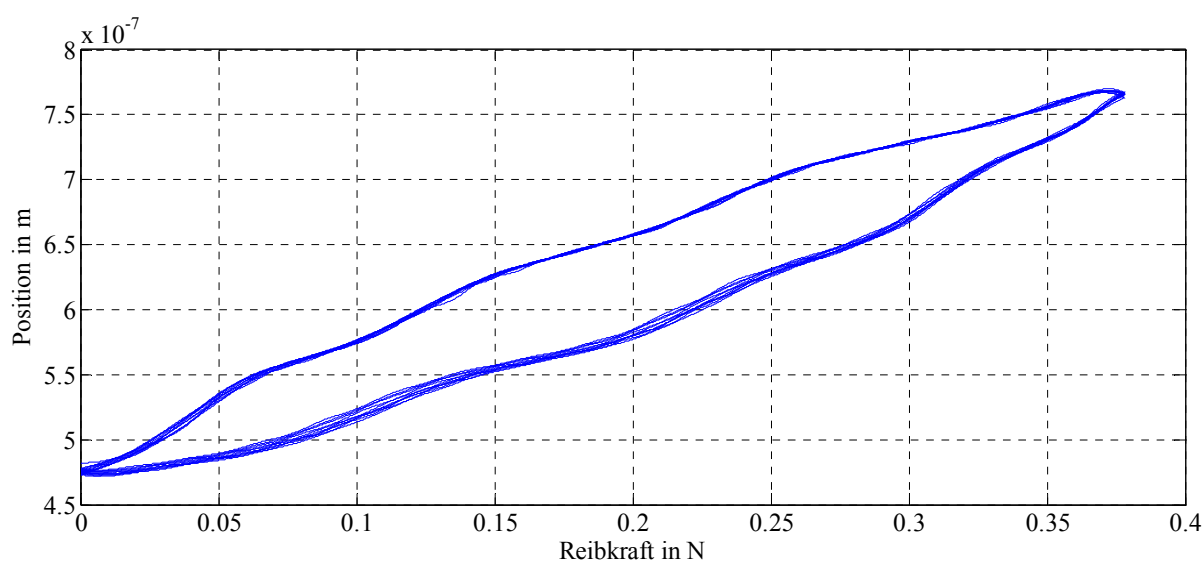


Abbildung 3.4: Reibkraft in Abhängigkeit der Position bei einer Bewegung im „Pre-Rolling“-Bereich

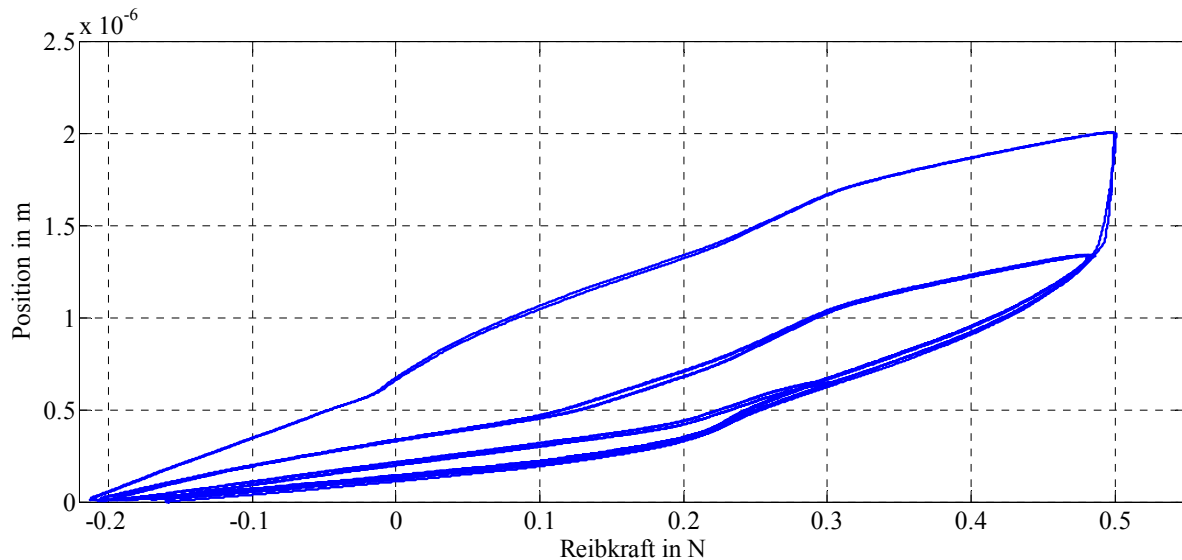


Abbildung 3.5: Reibkraft in Abhängigkeit der Position bei Bewegungen im „Pre-Rolling“-Bereich mit unterschiedlichen Amplituden

Hierzu werden die Antriebe der Y-Achse mit der Krafttrajektorie aus Abbildung A 5 beaufschlagt und die resultierende Systemreaktion beobachtet (siehe Abbildung A 6). Im Kraft-Positions-Diagramm sind beide Reibdomänen sehr gut sichtbar. Wirkt eine Kraft auf das System, die größer ist als die Losbrechkraft⁶, kommt es zu einer nur partiell reversiblen Bewegung. Wie Abbildung 3.6 zeigt, wird das System zu Beginn um circa 2100 nm ausgelenkt und bewegt sich nach Reduzierung der Antriebskraft um 600 nm in Richtung der Ausgangsposition zurück. Die verbleibende Auslenkung von 1500 nm ist nicht reversibel und kann daher als plastischer Bewegungsanteil angesehen werden.

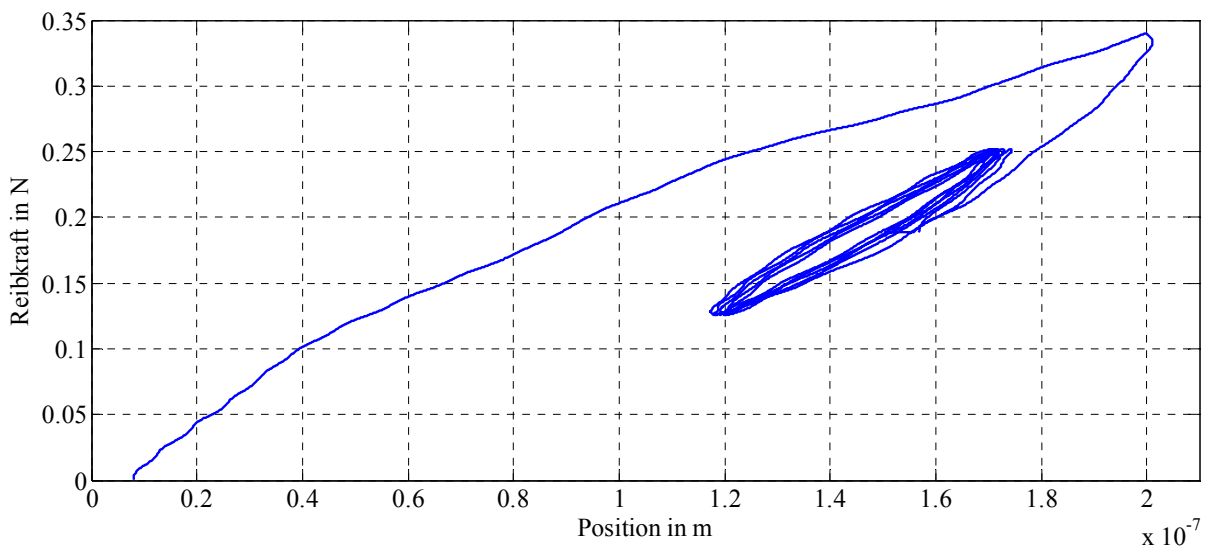


Abbildung 3.6: Reibkraft in Abhängigkeit der Position bei einer Bewegung mit reversiblen und irreversiblen Anteilen

⁶ In der Literatur wird die Losbrechkraft als „Break Away Force“ bezeichnet.

Im weiteren Verlauf des Graphen aus Abbildung A 5 wird das System mit einem sinusförmigen Kraftsignal unterhalb der Losbrechkraft angeregt und bleibt somit im rein elastischen Reibbereich. Auch dieses Verhalten ist gut in Abbildung 3.6 nachvollziehbar und entspricht dem aus Abbildung 3.5. Aus Sicht der Modellbildung ist der große und kontinuierliche Übergang zwischen den Reibbereichen das primäre Problem der Reibmodellierung. Des Weiteren konnte durch experimentelle Untersuchungen gezeigt werden, dass der Übergangsbereich besonders bei NPM-Systemen selten reproduziert werden kann und von einer Reihe weiterer Umgebungsvariablen beeinflusst wird [5, 6, 8, 20, 141]. Zu nennen wären in diesem Zusammenhang beispielsweise die Richtung der Bewegung, der Anstieg der angreifenden Kraft, die Verweildauer der Oberflächen im Ruhezustand, et cetera [12, 73].

Wie bereits erwähnt wurde, befindet sich das System im „Rolling“-Bereich, wenn alle Verbindungen zwischen den Kontaktflächen gerissen sind und die Körper sich irreversibel zu einander verschieben. Hier kommt es zu makroskopischen Bewegungen und der Betrag der Reibkraft ist im Wesentlichen nur noch abhängig von der Oberflächenbeschaffenheit, dem Oberflächenmaterial sowie dem Schmiermittel. Wird ein Schmierstoff verwendet, bildet sich ab einer stoffspezifischen Geschwindigkeit ein tragender Schmierfilm aus. Dieser Film sorgt für eine vollständige Trennung der reibenden Oberflächen und die Reibkraft nimmt daher in einem definierten Geschwindigkeitsbereich ab⁷ [12, 119]. Wird die Geschwindigkeit weiter gesteigert, führt dies wieder zu einem Anstieg der Reibkraft aufgrund sich ausbildender Turbulenzen im Schmierfilm. Ob das in der Literatur beschriebene Verhalten auch für NPM-Systeme gilt, wird ebenfalls anhand von experimentellen Untersuchungen gezeigt. Hierzu wird die Y-Achse des Demonstrators im geregelten Betrieb auf eine vorgegebene Geschwindigkeit beschleunigt und diese wird für einen definierten Zeitraum konstant gehalten. Während der Phase konstanter Geschwindigkeit gilt gemäß Gleichung (3.1) $F_A(t) = F_R(t)$. Somit entspricht die wirkende Reibkraft der Antriebskraft und eine Charakterisierung der Reibung bei konstanten Geschwindigkeiten ist möglich. Der Vorgang wird für eine endliche Anzahl von Geschwindigkeiten durchgeführt und es kann die Kennlinie in Abbildung 3.7 ermittelt werden. Die Bereiche nahe der Nullgeschwindigkeit⁸ sind zur Verdeutlichung im Anhang logarithmisch dargestellt (siehe Abbildung A 7 und Abbildung A 8). Es ist klar zu sehen, dass in den Führungen des xy-Demonstrators ein Schmiermittel verwendet wird, denn die Reibkraft vergrößert sich außerhalb der Nullgeschwindigkeit proportional zur Geschwindigkeit. Darüber hinaus kann aus den Messdaten entnommen werden, dass die verwendeten Lager keine Stribeck-Überhöhung aufweisen. Diese Beobachtung entspricht den Untersuchungen von De Moerlooze und Al Bender, welche das Reibverhalten von Präzisionskugellagern mit V-Nuten untersucht haben [2, 32]. Weiterhin kann eine klare Richtungsabhängigkeit der Reibkraft festgestellt werden, die auch ausführlich in der Literatur beschrieben wurde [12].

⁷ Das beschriebene Verhalten des anfänglichen Absinkens der Reibkraft wurde nach seinem Entdecker Stribeck-Überhöhung genannt.

⁸ In der vorliegenden Arbeit wird ein Bereich von ± 100 nm/s als Nullgeschwindigkeit bezeichnet.

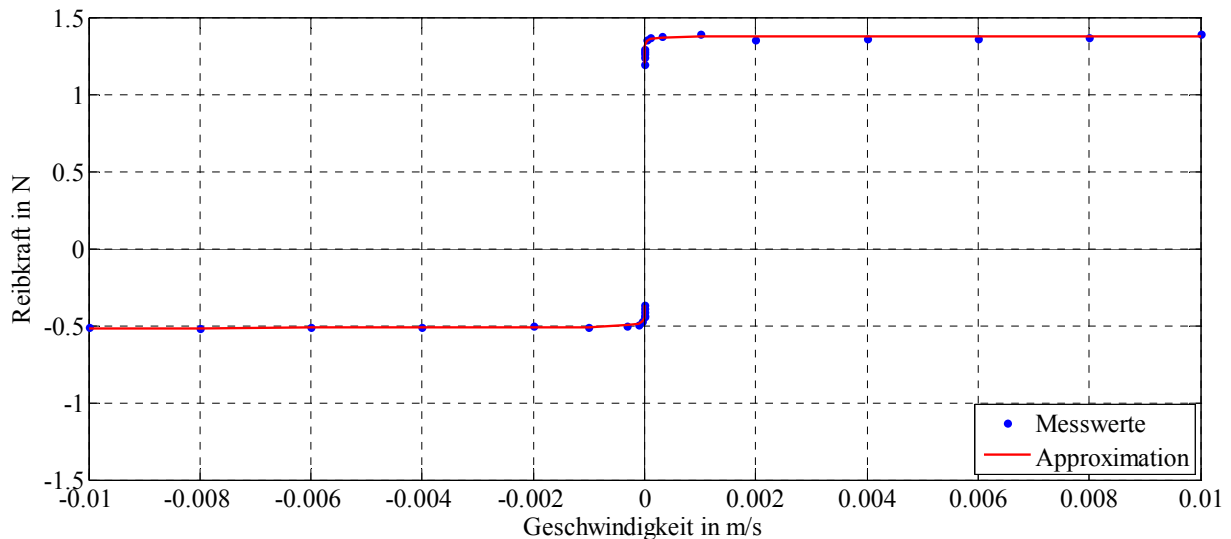


Abbildung 3.7: Statisches Reibverhalten bei makroskopischen Geschwindigkeiten

Neben den beschriebenen statischen Zusammenhängen haben im „Rolling“-Bereich auch dynamische Effekte einen nicht zu vernachlässigenden Einfluss. Zu nennen ist in diesem Zusammenhang beispielsweise der „frictional lag“-Effekt [13, 50, 90]. Dieses Phänomen beschreibt die Beobachtung, dass bei sich ändernden Geschwindigkeiten die Reibkraft erst mit einer systemspezifischen Verzögerung reagiert. Die Ursache für dieses Verhalten ist im Schmierfilm zu suchen, denn dessen Dicke repräsentiert einen Systemzustand, der sich nicht sprunghaft ändern kann. Um den Effekt am xy-Demonstrator nachzuweisen, wurde die Y-Achse im geregelten Betrieb mit einer periodisch auf- und abklingenden Geschwindigkeit (10 Hz) verfahren (siehe Abbildung A 9). Auch bei diesem Experiment können aufgrund der sehr kleinen Geschwindigkeiten die Trägheitskräfte vernachlässigt werden und es gilt $F_A(t) = F_R(t)$ (siehe Abbildung 3.3). In Abbildung A 10 ist die benötigte Antriebskraft dargestellt. Wird wie in Abbildung 3.7 die Reibkraft zur Geschwindigkeit aufgetragen, kommt es zur Ausbildung einer Hysterese, die umso ausgeprägter ist, je schneller sich die Geschwindigkeit ändert (siehe Abbildung 3.8).

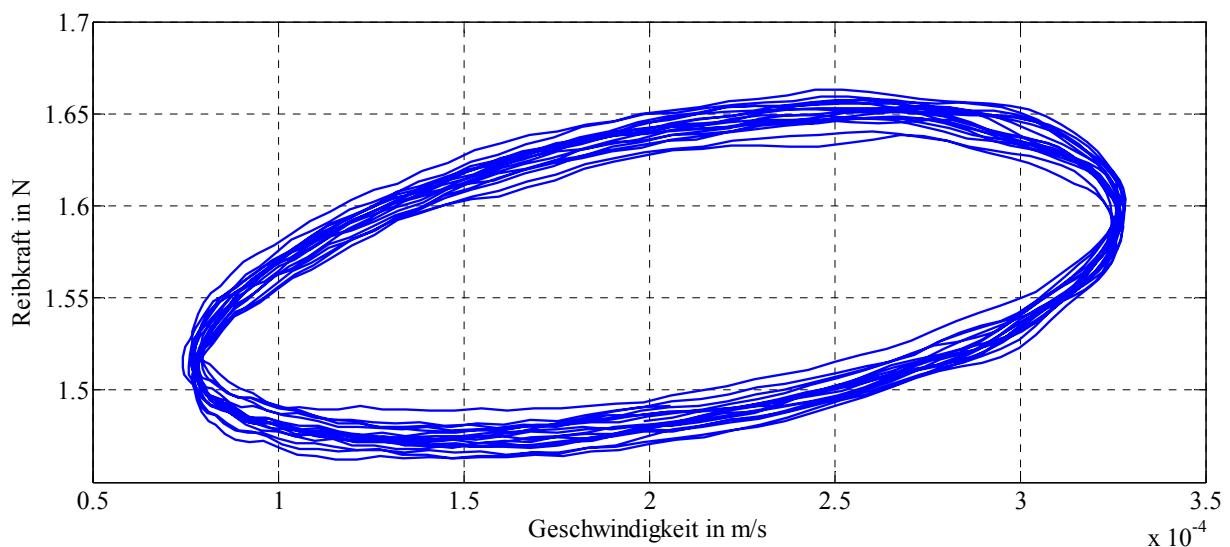


Abbildung 3.8: Reibkraft in Abhängigkeit der Geschwindigkeit (Frictional Lag)

Darüber hinaus ist eines der bekanntesten Reibphänomene der so genannte Stick-Slip oder Haft-Gleit-Effekt [12], welcher in Abbildung 3.9 dargestellt ist. Stick-Slip ist eine direkte Folge aus dem nichtlinearen Reibverhalten, denn die Reibkräfte im Bereich der Haftreibung, also im elastischen und elasto-plastischen Bereich, sind größer als bei rein plastischen Bewegungen. Bezogen auf den xy-Demonstrator bedeutet dies, dass die Kraft, die zur Überwindung der Haftreibung aufgewendet werden muss, größer ist als die Kraft, die zum Halten einer definierten Bewegungsgeschwindigkeit erforderlich ist. In Wechselwirkung mit Elastizitäten führt die beschriebene Charakteristik zum Stick-Slip-Effekt. Die Elastizität des Systems kann sowohl aus der mechanischen Bauteilelastizität als auch aus einem linearen Rückführgesetz resultieren. Somit tritt Stick-Slip häufig auf, wenn versucht wird, reibbehaftete Systeme mit klassischen linearen Regelungen zu beherrschen. Der durch den Regler verursachte nicht-lineare Grenzyklus lässt sich wie folgt beschreiben. Um die Haftreibung zu überwinden, wird eine relativ große Kraft auf die zu bewegende Masse ausgeübt. Dadurch beginnt der Körper zwar zu gleiten, aber die durch den Stelleingriff erzeugte Beschleunigung ist so stark, dass die resultierende Geschwindigkeit zu groß wird. Der lineare Regler reduziert daraufhin auf Basis des entstandenen Regelfehlers die Antriebskraft und der zu bewegende Körper wird wieder abgebremst. Aber auch hier ist die vom Regelungssystem erzeugte negative Beschleunigung zu groß und so wird die bewegte Masse zu stark abgebremst. Daraufhin bleibt der Körper wieder haften und der beschriebene Grenzyklus beginnt von Neuem [11, 93]. Das geschilderte Verhalten ist auch im Verlauf der Antriebskraft deutlich nachzuvollziehen. Wie in Abbildung A 11 zu erkennen ist, erzeugt der eingesetzte experimentell parametrisierte PID-Regler eine wellenförmige Stellgröße und verstärkt somit das Stick-Slip-Verhalten des Systems. Um diese Grenzyklen zu vermeiden, ist eine modellbasierte Regelung unabdingbar. Ein solcher Regelungsansatz basiert auf einem Systemmodell und somit wird eine einfach zu berechnende und gleichzeitig ausgesprochen detaillierte mathematische Beschreibung der Reibkraft benötigt. Aus diesem Grund widmen sich die folgenden Abschnitte einer Reibmodellierung, die es erlaubt, nahezu alle relevanten Reibphänomene nachzubilden.

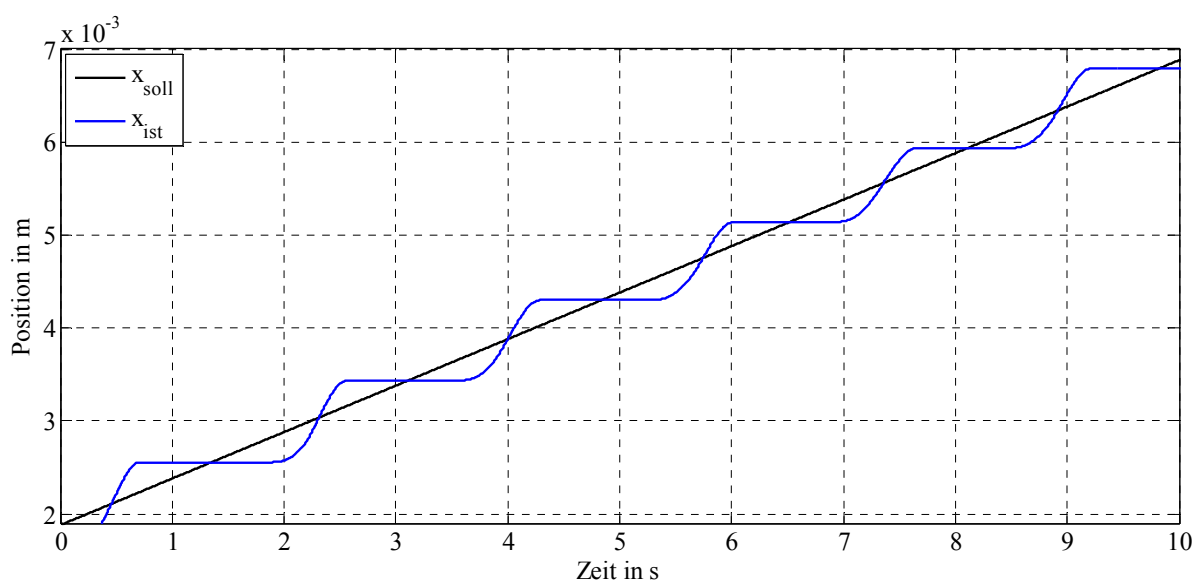


Abbildung 3.9: Stick-Slip-Effekt bei einer konstanten Geschwindigkeit von 0,5 mm/s

3.4.2 Stand der Technik auf dem Gebiet der Reibmodellierung

Das Phänomen der Reibung wurde zum ersten Mal von Leonardo da Vinci wissenschaftlich untersucht. Er stellte fest, dass die Reibkraft zwischen zwei Festkörpern entgegengesetzt der Bewegungsrichtung wirkt, proportional zur Belastung und unabhängig von der Größe der Auflagefläche ist [30]. Die beiden letzt genannten Eigenschaften der Reibung sind auch als die Amontons'schen Gesetze bekannt. Sie sind nach Guillaume Amontons benannt, der diese Gesetzmäßigkeiten im 17. Jahrhundert wiederentdeckte und veröffentlichte [4]. Darauf aufbauend wurden in den letzten Jahrhunderten eine Reihe von statischen wie dynamischen Reibmodellen publiziert, welche immer detaillierter versuchen, möglichst viele Reibphänomene zu beschreiben [12] (siehe Abschnitt 3.4.1).

Prinzipiell lassen sich die Versuche, Reibung zu modellieren in zwei übergeordnete Gruppen einteilen. Auf der einen Seite wären die so genannten „Blackbox“ Modelle anzuführen, die das stark nichtlineare Reibverhalten rein datengetrieben abbilden. Die zweite Modellklasse bilden die so genannten „Greybox“ oder physikalisch motivierten Reibmodelle, die durch eine analytische Beschreibung der Interaktion zwischen zwei rauen Oberflächen versuchen, die Reibung nachzubilden [135]. Im folgenden Abschnitt werden die bekanntesten „Greybox“ Modelle in chronologischer Reihenfolge dargestellt. Daran anknüpfend erfolgt ein kurzer Abriss über die bis dato eingesetzten datenbasierten Modelle.

3.4.2.1 Physikalisch motivierte Reibmodelle

Die physikalisch motivierten Reibmodelle können in statische und dynamische Reibmodelle eingeteilt werden. Als Vertreter der statischen Modelle können beispielhaft das klassische Coulomb- [95], das Stribeck- [119] oder das Karnopp-Modell [67] genannt werden.

Das Coulomb-Modell

Coulomb entwickelte im 18. Jahrhundert auf Basis der Erkenntnisse von da Vinci bzw. Amontons das nach ihm benannte Reibmodell. Das statische Coulomb-Modell beschreibt die Gleitreibung zwischen zwei Festkörpern auf Basis der Normalkraft $F_N(t)$, einem materialabhängigen Reibkoeffizienten μ und der Richtung der relativen Geschwindigkeit v zwischen diesen Körpern:

$$F_R = F_C \cdot \operatorname{sgn}(v) = \mu \cdot F_N \cdot \operatorname{sgn}(v) \quad \text{mit } v \neq 0 \quad (3.13)$$

Abbildung 3.10 a) zeigt exemplarisch den in Gleichung (3.13) definierten Zusammenhang. Weiterhin wurden unter anderem von Euler, Morin und Coulomb eine höhere Reibkraft beobachtet, wenn die relative Geschwindigkeit Null beträgt [95]. Um diesem Phänomen Rechnung zu tragen, wurde eine Unterscheidung zwischen Haftreibung und Gleitreibung eingeführt. Später entdeckte Reynolds, dass sich bei der Bewegung eines Körpers durch ein zähes Medium die Reibung mit zunehmender Geschwindigkeit erhöht [101].

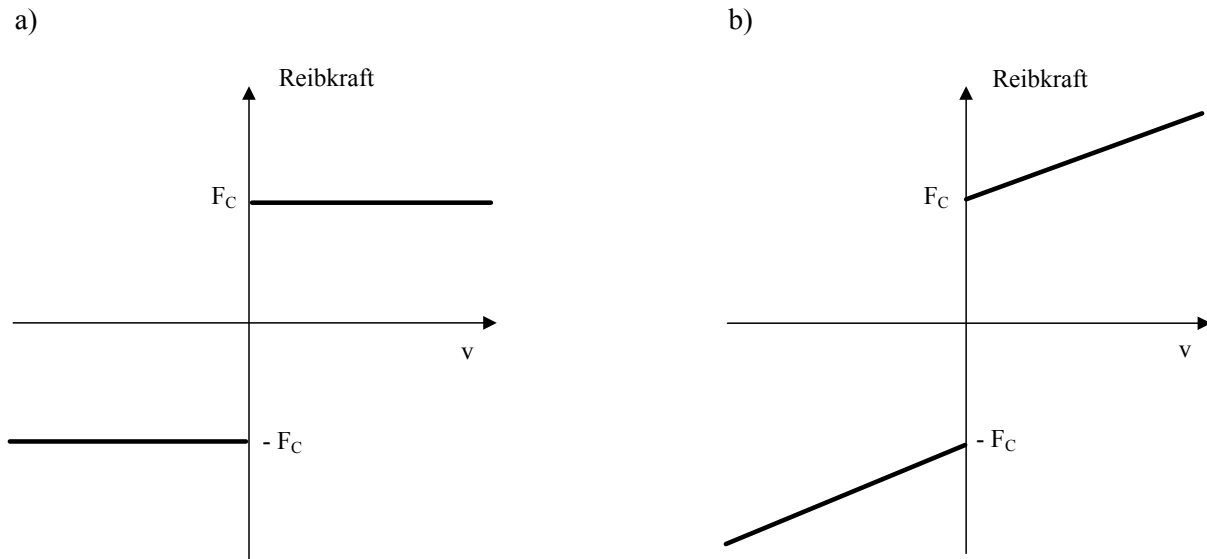


Abbildung 3.10: Coulomb-Reibung ohne (a) und mit (b) viskoser Reibung

Ursächlich hierfür ist die Ausbildung eines stabilen Schmierfilmes zwischen den reibenden Oberflächen und ab diesem Punkt dominieren hydrodynamische Effekte das Reibverhalten. Das resultierende Verhalten wird als viskose Reibung oder Flüssigkeitsreibung bezeichnet, und somit kann das klassische Coulomb-Modell um einen viskosen Anteil erweitert werden (siehe Abbildung 3.10 b)):

$$F_R = F_C \cdot \operatorname{sgn}(v) + \sigma_2 \cdot v \quad \text{mit } v \neq 0 \quad (3.14)$$

Das Stribeck-Modell

Im Zuge der Erweiterung des viskosen Coulomb-Modells stellte Stribeck zu Beginn des 20. Jahrhunderts ein nach ihm benanntes Reibmodell vor (siehe Abbildung 3.11 a)). In diesem statischen Modell wird erstmalig der Effekt berücksichtigt, dass die Reibung beim Übergang zwischen Haft- und Gleitreibung höher ist als bei kleinen Geschwindigkeiten. Wie Abbildung 3.11 b) zeigt, lässt sich die Stribeckkurve in Abhängigkeit der relativen Geschwindigkeit in drei Bereiche einteilen. Setzt nach der Überwindung der Haftreibungskraft F_S eine Bewegung zwischen den Kontaktflächen ein, baut sich langsam ein Schmierfilm auf. Zu Beginn trennt dieser nur teilweise die reibenden Flächen voneinander und es wird von Festkörper- bzw. Grenzreibung gesprochen (Bereich I). Mit zunehmender Geschwindigkeit bildet sich anschließend ein dünner Schmierfilm und die Oberflächen berühren sich nur noch punktuell. Es dominiert die so genannte Mischreibung (Bereich II). Wird die relative Geschwindigkeit weiter erhöht, nimmt die Dicke des Schmierfilmes zu und trennt ab einem systemspezifischen Punkt vollständig die interagierenden Oberflächen voneinander (Bereich III). In der Literatur findet sich eine Reihe von Gleichungen zur Beschreibung der Stribeckkurve. Beispielhaft kann das Modell von Stribeck wie folgt beschrieben werden:

$$F_R = \left(F_C + (F_S - F_C) e^{-|v/v_s|^2} \right) \cdot \operatorname{sgn}(v) + \sigma_2 \cdot v \quad \text{mit } v \neq 0 \quad (3.15)$$

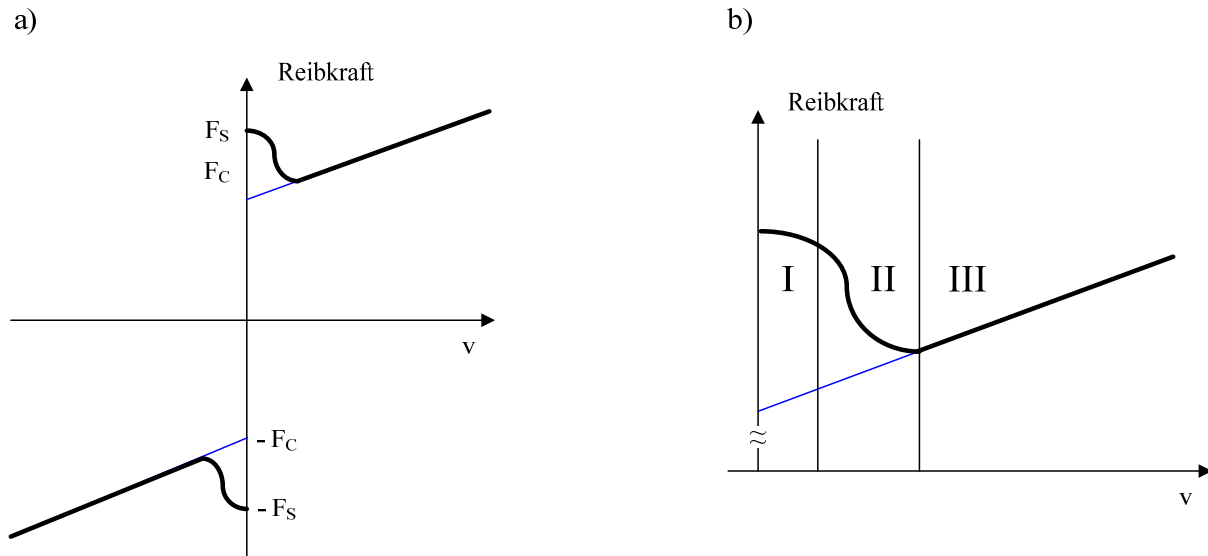


Abbildung 3.11: Stribeckkurve (a) und Detailausschnitt der Stribeck-Überhöhung (b)

Der Parameter v_S in Gleichung (3.15) wird als Stribeck-Geschwindigkeit bezeichnet und bestimmt die Form des Überganges zwischen Haft- und Gleitreibung. Sowohl das Coulomb- als auch das Stribeck-Modell sind rechentechnisch nur bedingt umsetzbar und somit im Sinne einer späteren Verwendung als Bestandteil einer Regelung ungeeignet. Problematisch ist hierbei im Besonderen der Umschaltzeitpunkt der *Sgn*-Funktion.

Das Karnopp-Modell

Zur Behebung der Diskontinuität der *Sgn*-Funktion wurden eine Vielzahl von Lösungen vorgestellt. Ein genereller Ansatz ist die Approximation der *Sgn*-Funktion mit einer sehr steil ansteigenden Funktion wie etwa einem Tangens Hyperbolicus. Nachteilig wirkt sich bei dieser Lösung aber aus, dass eine numerische Integration nahe der Nullgeschwindigkeit aufgrund des extrem steilen Anstieges nicht mehr echtzeitfähig durchführbar ist. Darüber hinaus bildet ein derartiges Modell den Haftbereich nicht ab. Eine gänzlich andere Lösung des Problems wurde von Karnopp vorgeschlagen, indem ein Intervall $[-D_v, +D_v]$ um $v = 0$ eingeführt wurde. In diesem Bereich ist die Reibung abhängig von der angreifenden Kraft F und das Modell kann wie folgt formuliert werden:

$$F_R(v, F) = \begin{cases} -F_C \cdot \operatorname{sgn}(v) & |v| > D_v \\ -\operatorname{sgn}(F) \cdot \max(F, F_H) & |v| \leq D_v \end{cases} \quad (3.16)$$

Die bisher vorgestellten statischen Reibmodelle sind zwar sehr leicht berechenbar, bilden aber nur einen statischen Zusammenhang zwischen wirkender Reibkraft und einer konstanten relativen Bewegungsgeschwindigkeit ab. Des Weiteren können diese Modelle die Reibkraft bei Geschwindigkeiten nahe Null strukturell nicht nachbilden und eignen sich daher nicht für die Modellierung der Reibung im Nanometerbereich.

Aus diesem Grund müssen besonders mit Blick auf NPM-Systeme die dynamischen Effekte der Reibung (siehe Abschnitt 3.4.1) bei der Modellierung mit berücksichtigt werden. Dies gelingt mit dynamischen Reibmodellen, welche Ende der 60'er Jahre erstmals in der Rege-

lungstechnik angewendet wurden. Im Folgenden werden die bekanntesten dynamischen Reibmodelle in chronologischer Reihenfolge skizziert.

Das Dahl-Modell

Das dynamische Reibverhalten von gelagerten und ungelagerten Bewegungen wurde zum ersten Mal von Dahl in einem gleichnamigen Reibmodell mathematisch erfasst [31]. Im Zuge ausgiebiger experimenteller Untersuchungen an Servoantrieben stellte Dahl bei gelagerter und ungelagerter Reibung ein ähnliches dynamisches Verhalten fest und beschrieb dieses mithilfe einer abgewandelten Spannungs-Dehnungs-Kurve aus der klassischen Festkörpermechanik:

$$\dot{F}_R(t) = \sigma_s \left(1 - \frac{F_R(t)}{F_C} \cdot \operatorname{sgn}(\dot{x}(t)) \right)^\beta \cdot \dot{x}(t) \quad (3.17)$$

$F_R(t)$ repräsentiert in Gleichung (3.17) die dynamische Reibkraft, σ_s einen Steifigkeitskoeffizienten und β einen Gestaltparameter. Dieses vergleichsweise einfache Modell stellt eine Generalisierung des statischen Coulomb-Modells dar und ist in der Lage, einige Reibphänomene im elastischen Bereich nachzubilden. Effekte, wie die Stribeck-Überhöhung oder das Reibgedächtnis (siehe Abschnitt 3.4.1) werden von diesem Modell nicht berücksichtigt.

Das „Lund-Grenoble“ (LuGre) -Reibmodell

Eine deutliche Verbesserung bezüglich der Exaktheit dynamischer Reibmodelle gelang Carlos Canudas de Wit *et al.* bei der Formulierung des „Lund-Grenoble“-Modells [33, 35, 90]. Bei der dynamischen Modellierung der Reibkraft wurde der Vorschlag von Haessing und Friedland [47] aufgegriffen, die Oberflächenstruktur zweier reibender Körper mit einer endlichen Anzahl von Borsten zu approximieren. Schematisch ist dieser Zusammenhang in Abbildung 3.12 dargestellt. Diese Borsten verhalten sich wie elastische Federn, und die wirkende Reibkraft wird als die Kraft angesehen, welche notwendig ist, um die Borsten zu deformieren. Falls die angreifende Kraft groß genug ist, erreichen die Borsten ihre maximal mögliche Auslenkung und beginnen übereinander zu gleiten. Die Durchbiegung einer Borste wird mit einer Zustandsvariablen $z(t)$ beschrieben. Um die Modellkomplexität zu begrenzen, wird angenommen, dass sich nur auf einer der beiden Oberflächen elastische Borsten befinden und der Durchbiegungszustand $z(t)$ für alle elastischen Borsten identisch ist.

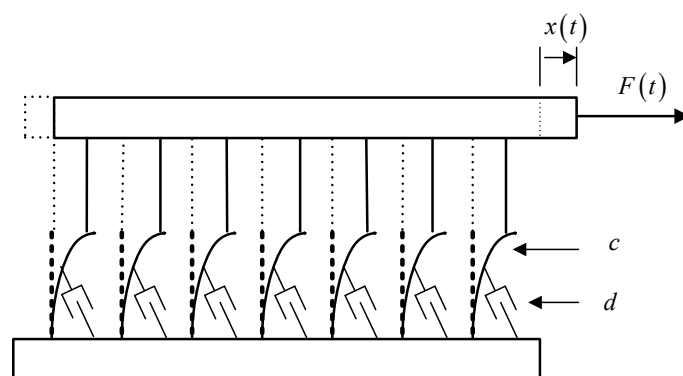


Abbildung 3.12: Modellvorstellung des so genannten Borstenansatzes

Somit ergibt sich ein leicht berechenbares dynamisches Modell, welches die wesentlichen Aspekte der Reibung abbildet:

$$F_R(t) = \sigma_0 \cdot z(t) + \sigma_1 \cdot \dot{z}(t) + \sigma_2 \cdot \dot{x}(t) \quad (3.18)$$

$$\dot{z}(t) = \dot{x}(t) - \frac{|\dot{x}(t)|}{g(\dot{x}(t))} \cdot z(t) \quad (3.19)$$

$$g(\dot{x}(t)) = \frac{1}{\sigma_0} \cdot F_{ss}(\dot{x}(t)) \quad (3.20)$$

$$F_{ss}(\dot{x}(t)) = F_C + (F_S - F_C) e^{-(\dot{x}(t)/v_s)^\beta} \quad (3.21)$$

Nach Gleichung (3.18) setzt sich die Reibkraft im „LuGre“-Modell aus drei Komponenten zusammen. Die ersten zwei Summanden repräsentieren die Kraft, welche aufgebracht werden muss, um die Borsten zu deformieren. Der dritte Summand bildet den viskosen, also geschwindigkeitsproportionalen Teil der Reibung ab und σ_2 spiegelt hierbei die Viskosität des verwendeten Gleitmittels wider. Der innere Zustand $z(t)$ wird in Gleichung (3.19) berechnet und basiert auf der aktuellen Geschwindigkeit $\dot{x}(t)$ als Eingangsgröße. Der Term $F_{ss}(\dot{x}(t))$ steht hier für die Stribeckkurve, welche in den Gleichungen (3.20) und (3.21) mittels einer Exponentialfunktion aus statischer Coulombkraft F_C , Haftreibungskraft F_S sowie Stribeck-Geschwindigkeit v_s berechnet wird. Der Parameter σ_0 beschreibt in diesem Zusammenhang die mikro-elastische Steifigkeit der Borsten. Dieses vergleichsweise einfache Modell ist rechenstechnisch leicht zu implementieren und bildet mit nur einer Zustandsvariablen die Mehrheit der Reibeffekte im elastischen und plastischen Reibbereich ab. Aus diesem Grund wurde das „LuGre“-Modell seit seiner Veröffentlichung mit durchweg großem Erfolg eingesetzt. Nichtsdestotrotz besitzt das „LuGre“-Modell einige Schwächen, denn örtliche Auslenkungen besitzen unabhängig von ihrer Größe immer einen reversiblen und irreversiblen Anteil. Dieses Verhalten äußert sich zum Beispiel in der Drift des Modells bei einer sinusförmigen Anregung unterhalb der Haftreibungskraft F_S .

Das elasto-plastische Reibmodell

Um die Unzulänglichkeiten des „LuGre“-Modells zu beseitigen, stellten Dupont *et al.* 2002 eine Erweiterung des Modells vor [39, 40]. Die Berechnung des Zustandes $z(t)$ wurde stark modifiziert und so wird es möglich, den elastischen Reibbereich ohne plastische Bewegungsanteile zu modellieren. Auch wird durch diese Erweiterung der Übergangsbereich zwischen elastischer und plastischer Reibdomäne detaillierter beschrieben. Es wird davon ausgegangen, dass sich die Verschiebung $x(t)$ eines Körpers aus zwei Komponenten zusammensetzt: Die mittlere Auslenkung der Borsten $z(t)$ stellt die elastische (reversible) Komponente dar. Der zweite Teil $w(t)$ ist gleich der plastischen (irreversiblen) Verschiebung des Körpers, die auftritt, wenn die Borsten übereinander gleiten. Es gilt:

$$x(t) = z(t) + w(t) \quad (3.22)$$

Weiterhin wird bei der Bewegung eines Körpers zwischen elastischer, elasto-plastischer und plastischer Bewegung unterschieden. Bis zu einer kritischen Auslenkung z_{ba} gilt $\dot{z}(t) = \dot{x}(t)$

und $\dot{w}(t) = 0$. Es liegt also eine rein elastische Bewegung vor. Beträgt die Borstenauslenkung mehr als z_{ba} , aber weniger als die maximale Biegung der Borsten $z_{ss}(\dot{x}(t))$, wird die Verschiebung dem elasto-plastischen Bereich zugeordnet, da nun elastische und plastische Bewegungsanteile vorhanden sind. Dieser Übergangsbereich kann mit $\dot{x}(t) = \dot{z}(t) + \dot{w}(t)$ beschrieben werden. Erreicht die Auslenkung der Borsten den maximalen Wert $z_{ss}(\dot{x}(t))$, kommt es zu keiner weiteren Biegung der Borsten und diese beginnen übereinander zu gleiten. Die Bewegung ist nun rein plastisch und es gilt $\dot{z}(t) = 0$ und $\dot{x}(t) = \dot{w}(t)$. Dieses Verhalten kann wie folgt mathematisch formuliert werden:

$$F_R(t) = \sigma_0 \cdot z(t) + \sigma_1 \cdot \dot{z}(t) + \sigma_2 \cdot \dot{x}(t) \quad \text{mit } \sigma_0, \sigma_1, \sigma_2 > 0 \quad (3.23)$$

$$\dot{z}(t) = \dot{x}(t) \cdot \left(1 - \alpha(z(t), \dot{x}(t)) \frac{z(t)}{z_{ss}(\dot{x}(t))} \right) \quad (3.24)$$

$$\alpha(z(t), \dot{x}(t)) = \begin{cases} 0, & |z| \leq z_{ba} \\ 0 < \alpha_m(z(t)) < 1, & z_{ba} < |z| < z_{ss}(\dot{x}(t)) \cap \text{sgn}(\dot{x}(t)) = \text{sgn}(z(t)) \\ 1, & |z| < z_{ss}(\dot{x}(t)) \cap \text{sgn}(\dot{x}(t)) = \text{sgn}(z(t)) \\ 0, & \text{sgn}(\dot{x}(t)) \neq \text{sgn}(z(t)) \end{cases} \quad (3.25)$$

$$\text{mit } 0 < z_{ba} < z_{ss}(\dot{x}(t))$$

$$\alpha_m(z(t)) = \frac{1}{2} \sin \left(\pi \cdot \frac{z(t) - \frac{z_{ss}(\dot{x}(t)) - z_{ba}}{2}}{z_{ss}(\dot{x}(t)) - z_{ba}} \right) + \frac{1}{2} \quad \text{mit } z_{ba} < |z| < z_{ss}(\dot{x}(t)) \quad (3.26)$$

$$z_{ss}(\dot{x}(t)) = \frac{\text{sgn}(\dot{x}(t))}{\sigma_0} \cdot F_{ss}(\dot{x}(t)) \quad (3.27)$$

$$F_{ss}(\dot{x}(t)) = F_C + (F_S - F_C) e^{-(\dot{x}(t)/v_s)^2} \quad (3.28)$$

Das in den Gleichungen (3.23) bis (3.28) definierte Modell bildet eine Vielzahl der Reibeffekte ab, die bei Kugellagerführungen im Nanometerbereich beobachtet werden können [140] (siehe Abschnitt 3.4.1). Dazu gehört die Nachbildung der elastischen und plastischen Reibdomäne sowie des Übergangs zwischen diesen Bereichen. Weiterhin ist das Modell in der Lage, den Stick-Slip-Effekt zu reproduzieren.

Das „Generalized Maxwell-Slip“ (GMS) -Reibmodell

Lampaert *et al.* lösten sich Ende der 90'er Jahre von der Idee, die Reibung mithilfe des Borstenansatzes zu beschreiben. Sie modellieren die Reibung durch eine endliche Anzahl von parallel angeordneten, masselosen und über Federn verbundenen Körpern, so genannten „Maxwell-Slip“-Elementen. Vorteil dieses „Basic-Maxwell-Slip“-Ansatzes ist dessen Fähigkeit, das Hystereseverhalten der Reibkraft im elastischen Bereich fehlerfrei abzubilden. Der älteste Vertreter dieser Gruppe ist das so genannte „Leuven“-Modell [74, 75, 120], welches jedoch erhebliche Probleme bei der rechentechnischen Umsetzung bereitet. Dieses Manko wurde

kurze Zeit später mit der Vorstellung des GMS-Modells behoben, indem die Berechnung der inneren Zustände des „Leuven“-Modells stark abgewandelt wurde [3, 72, 126]. Prinzipiell basiert das GMS-Modell auf der in Abbildung 3.13 dargestellten Vorstellung, dass M parallel angeordnete Elemente durch die Bewegung der reibenden Körper angeregt werden. Das Verhalten jedes dieser Elemente ist bestimmt durch eine Federkonstante k_i und durch einen Positionsschwellwert Δ_i . Der Schwellwert bestimmt, wann das entsprechende Element haftet und wann es gleitet. Haftet ein Element, so wird die Feder durch die erregende Bewegung gedehnt und das Element leistet einen variablen Kraftbeitrag $F_i(t)$. Ist der Positionsschwellwert überschritten, beginnt das Element schrittweise zu gleiten und leistet durch die nun konstante Federdehnung $\delta_i(t)$ einen konstanten Kraftbeitrag. Die Einzelkräfte $F_i(t)$ werden aufsummiert und ergeben die resultierende Reibkraft. Die Besonderheit des GMS-Modells ist nun, dass die Schwellwerte in Abhängigkeit von der Stribeckkurve berechnet werden und somit mit der Geschwindigkeit variieren. Gleichung (3.29) bis (3.31) stellen die GMS-Modellstruktur dar:

$$F_R(t) = \sum_{i=1}^M \left[k_i \cdot \delta_i(t) + \sigma_i \cdot \dot{\delta}_i(t) \right] + f_{cn}(\dot{x}(t)) \quad (3.29)$$

$$\dot{\delta}_i(t) = \dot{x}(t) \quad \text{wenn } \delta_i < g_i(\dot{x}(t)) \quad (3.30)$$

$$\dot{\delta}_i(t) = \text{sgn}(\dot{x}(t)) \cdot C_i \cdot \left(1 - \frac{\delta_i(t)}{g_i(\dot{x}(t))} \right) \quad \text{wenn } \delta_i \geq g_i(\dot{x}(t)) \quad (3.31)$$

Der Term $f_{cn}(\dot{x}(t))$ wird hier definiert mit $k_v \cdot \dot{x}(t)$ und spiegelt den viskosen Teil der Reibkraft wider. Die Funktion $g_i(\dot{x}(t))$ wird nach Gleichung (3.20) berechnet und basiert auf der Stribeckkurve. C_i ist ein Attraktionsparameter, der die Stribeckkurve skaliert und σ_i ein zusätzlicher Freiheitsgrad, der das visko-elastische Verhalten des betrachteten Elementes bestimmt. Um die Modellkomplexität zu reduzieren, wird sowohl das Stribeckverhalten $g_i(\dot{x}(t))$ wie auch der Modellparameter C_i für alle „Maxwell-Slip“-Elemente als gleich angenommen.

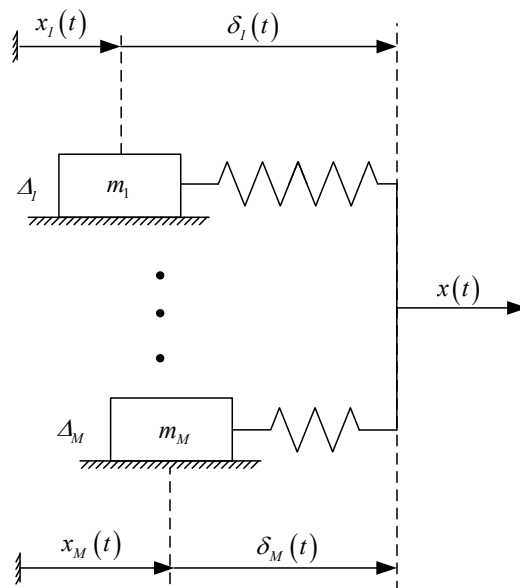


Abbildung 3.13: „Basic-Maxwell-Slip“ Modellvorstellung

Ist die Modellgüte in dieser Ausprägung nicht ausreichend, kann ein Skalierungsfaktor γ_i eingeführt werden, der die Anzahl der Freiheitsgrade wieder vergrößert [3]. Somit ergeben sich die zusätzlichen Modellgleichungen $g_i(\dot{x}(t)) = \gamma_i \cdot g(\dot{x}(t))$ und $C_i = \gamma_i \cdot C$. Das GMS-Modell liefert sehr gute Reibkraftvorhersagen und ist in der Lage, die ortsunabhängige Hysterese im elastischen Reibbereich abzubilden [73] (siehe Abschnitt 3.4.1). Weiterhin besitzt es aber wie alle bisher vorgestellten dynamischen Reibmodelle den Nachteil, dass die Parameteridentifikation einen nicht unerheblichen experimentellen Aufwand erfordert.

Die “Dynamic NonLinear Regression with direct application of eXcitation” (DNLRX)-Identifikationsmethode

Rizos *et al.* entwickelten, basierend auf der in Abbildung 3.13 dargestellten “Basic-Maxwell-Slip“ Modellvorstellung, einen zum GMS-Modell alternativen Ansatz, welcher den Aufwand zur Identifikation der Modellparameter minimiert [102, 103, 104]. Darüber hinaus bildet das DNLRX-Modell im Gegensatz zu den bisher vorgestellten Modellen nicht nur die Reibung sondern auch die Trägheit des kinematischen Systems ab.

Es wird ein einfaches mechanisches System mit Reibung betrachtet (siehe Abbildung 3.14). Dieses besteht aus einer Masse m , einer linearen Feder mit der Federkonstante c und einem Dämpfer mit einer Dämpfungskonstante d . Das System wird angeregt durch eine Kraft $F_A(t)$ und die (nicht messbare) Reibkraft $F_R(t)$ wirkt der resultierenden Bewegung entgegen.

Das dynamische Verhalten des Systems kann wie folgt beschrieben werden:

$$m \cdot \ddot{x}(t) + d \cdot \dot{x}(t) + c \cdot x(t) = F_A(t) - F_R(t) \quad (3.32)$$

$\ddot{x}(t)$ stellt die Beschleunigung $\dot{x}(t)$ die Geschwindigkeit und $x(t)$ die resultierende Position dar. Im Diskreten kann die benötigte Beschleunigung und Geschwindigkeit mithilfe einer gleitenden Mittelwertdarstellung der Ordnung n_v bzw. n_a und den Koeffizienten p_i bzw. q_i approximiert werden:

$$\dot{x}(k) = \sum_{i=0}^{n_v} p_i \cdot x(k-i) \quad \text{mit } t = k \cdot T \quad (3.33)$$

$$\ddot{x}(k) = \sum_{i=0}^{n_a} q_i \cdot x(k-i) \quad \text{mit } t = k \cdot T \quad (3.34)$$

T ist hierbei die Tasterperiode. Die zeitdiskrete Reibkraft $F_R(t)$ wird mit Hilfe eines zeitdiskreten „Basic-Maxwell-Slip“-Ansatzes modelliert.

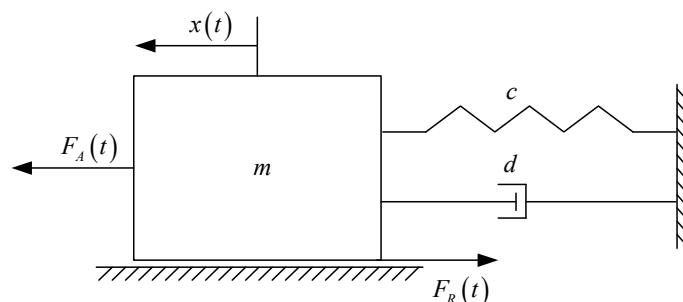


Abbildung 3.14: Einfaches System mit Reibung

Somit können die Federauslenkungen eines betrachteten Elementes analog zu den Gleichungen (3.30) und (3.31) bestimmt werden:

$$\delta_i(k+i) = \text{sgn}[x(k+1) - x(k) + \delta_i(k)] \cdot \min\{|x(k+1) - x(k) + \delta_i(k)|, \Delta_i\} \quad (3.35)$$

Im Unterschied zum GMS-Modell sind die Positionsschwellwerte Δ_i der “Maxwell-Slip”-Elemente konstant. Darüber hinaus wird beim DNLRX-Modell die Reibkraft nicht nur durch die einzelnen “Maxwell-Slip”-Elemente erzeugt, sondern es wird zusätzlich eine Abhängigkeit der Reibung von den vergangenen Positionen zugelassen. Grund hierfür ist eine Einführung zusätzlicher Freiheitsgrade, um unmodellierten Anteilen in der Systemdynamik Rechnung zu tragen. Somit kann die diskretisierte Reibkraft wie folgt berechnet werden:

$$F_R(k) = \sum_{i=0}^{n_r} r_i \cdot x(k-i) + \sum_{i=0}^{n_\delta} \Theta_i^T \cdot \delta(k-i) \quad (3.36)$$

mit

$$\delta(k) = [\delta_1(k) \quad \dots \quad \delta_M(k)]^T$$

$$\Theta_i = [\theta_{1,i} \quad \dots \quad \theta_{M,i}]^T$$

Diese Schreibweise lässt den Schluss zu, dass die Reibung berechnet werden kann, indem die Positionswerte mit einem FIR-Filter der Ordnung n_r (mit den Koeffizienten r_j) und die Federauslenkungen mit einem M -dimensionalen FIR-Filter der Ordnung n_δ (mit den Vektorkoeffizienten Θ_j) gefiltert werden. Die Gleichungen (3.33), (3.34) und (3.36) liefern Schätzwerte für die Geschwindigkeit, die Beschleunigung und die Reibkraft in Abhängigkeit der vergangenen Positionswerte. Werden diese Approximationen in die diskrete Form der Gleichung (3.32) eingesetzt, entsteht das folgende Systemmodell:

$$m \cdot \sum_{i=0}^{n_a} q_i \cdot x(k-i) + d \cdot \sum_{i=0}^{n_v} p_i \cdot x(k-i) + c \cdot x(k) = F_A(k) - \left[\sum_{i=0}^{n_r} r_i \cdot x(k-i) + \sum_{i=0}^{n_\delta} \Theta_i^T \cdot \delta(k-i) \right] \quad (3.37)$$

Dieses Modell kann nach der erregenden Kraft umgestellt werden und so ergibt sich eine inverse Modellbeschreibung des dynamischen Systemverhaltens:

$$F_A(k) = m \cdot \sum_{i=0}^{n_a} q_i \cdot x(k-i) + d \cdot \sum_{i=0}^{n_v} p_i \cdot x(k-i) + c \cdot x(k) + \sum_{i=0}^{n_r} r_i \cdot x(k-i) + \sum_{i=0}^{n_\delta} \Theta_i^T \cdot \delta(k-i) \quad (3.38)$$

Im letzten Schritt werden alle Terme, welche von der Position $x(t)$ abhängig sind, zusammengefasst:

$$F_A(k) = \sum_{i=0}^{n_x} g_i \cdot x(k-i) + \sum_{i=0}^{n_\delta} \Theta_i^T \cdot \delta(k-i) \quad \text{mit } n_x = \max\{n_a, n_v, n_r\} \quad (3.39)$$

In den Faktoren g_i werden die Systemparameter mit $m \cdot q_i$, c , $d \cdot p_i$, r_i gebündelt und es gilt beispielsweise $g_0 = m \cdot q_0 + d \cdot p_0 + c + r_0$. Diese Art der Darstellung ermöglicht eine Modellierung des inversen Systemverhaltens lediglich auf Basis von zwei FIR-Filtern.

3.4.2.2 Datenbasierte Reibmodelle

Neben diesen physikalisch motivierten Reibmodellen kamen mit Einführung der digitalen Regelungstechnik verstärkt datengetriebene „Blackbox“-Modelle auf. Als die populärsten Vertreter wären hier die künstlichen neuronalen Netze (KNN) zu nennen, denn diese eignen sich ausgesprochen gut zur Modellierung starker Nichtlinearitäten [135]. In der Vergangenheit wurden angepasste Netzstrukturen bereits in verschiedenen Systemen zur Reibkraftvorhersage eingesetzt [114, 135]. Als geeignet stellten sich in diesem Zusammenhang insbesondere rekurrente [85] sowie inhibitorische [15] Netzstrukturen heraus. Darüber hinaus konnte in [129] nachgewiesen werden, dass auch herkömmliche Feedforward-Netztopologien in der Lage sind, das inverse Systemverhalten von NPM-Systemen zufriedenstellend nachzubilden. Unter gewissen Randbedingungen können auch Fuzzy-basierte Beobachterkonzepte als eine Art der datengetriebenen Modellierung angesehen werden. Daher wurde bis dato bei einer Reibkraftmodellierung auch dieses Konzept berücksichtigt, denn aufgrund der Struktur von Fuzzy-Systemen sind diese prädestiniert für eine modellbasierte Nachbildung von extremen Nichtlinearitäten [54, 87].

Da die Modellierung der Reibkraft mithilfe von datengetriebenen Modellen aufgrund der systemabhängigen Spezialisierung nur sehr schwer generalisiert darstellbar ist, soll im Folgenden lediglich die in [129] vorgestellte Lösung als Beispiel näher beleuchtet werden.

Vollvermaschtes Feedforward (FF) -Netzwerk

Mit dem Ziel einer für NPM-Systeme angemessenen Reibkraftmodellierung wurde im Rahmen dieser Arbeit ein vollvermaschtes FF-Netz eingesetzt. Wie die Abbildung 3.15 zeigt, besitzt das Netz nur eine Hiddenschicht und die Zahl der Eingangsneuronen ist frei wählbar. Vorteil der Feedforward-Netzstruktur ist die strukturbedingte schnelle Berechenbarkeit des Netzausganges. Darüber hinaus kann die Berechnung gut parallelisiert werden, da jede Ausgabe eines Neurons nur von der davorliegenden Schicht abhängig ist. Die Dauer der Berechnung korreliert somit weitestgehend mit der Anzahl der Schichten und nicht mit der Anzahl der Neuronen. Nachteil der verwendeten Netzstruktur ist jedoch, dass das Netz aufgrund fehlender interner Rückkopplungen kein Gedächtnis besitzt. Um trotzdem dynamisches Verhalten abbilden zu können, werden dem Netz neben dem aktuellen zusätzlich verzögerte Eingangswerte übergeben und somit „künstlich“ ein Gedächtnis geschaffen. Wie Abbildung 3.15 zeigt, setzt sich der Eingang des KNN aus zwei Vektoren zusammen. Der erste Vektor U_1 besteht aus dem aktuellen und $n - 1$ vergangenen Positionswerten. Der zweite Vektor U_2 beinhaltet den aktuellen sowie $m - 1$ vergangene Geschwindigkeitswerte. In diskreter Darstellung ergibt sich somit:

$$\mathbf{U}_1(k) = \begin{pmatrix} x(k) \\ x(k-1) \\ \vdots \\ x(k-n) \\ x(k-n+1) \end{pmatrix}, \quad \mathbf{U}_2(k) = \begin{pmatrix} \dot{x}(k) \\ \dot{x}(k-1) \\ \vdots \\ \dot{x}(k-m) \\ \dot{x}(k-m+1) \end{pmatrix} \quad (3.40)$$

Die Länge des Gedächtnisses des Netzes ist somit durch die Parameter n und m einstellbar und muss in Abhängigkeit der Zeitkonstanten des Systems gewählt werden. Die Werte für Position und Geschwindigkeit werden auf das Intervall $[0, 1]$ normiert. Diese Normierung ist üblich, da die gewählten Aktivierungsfunktionen in der Hiddenschicht nur in diesem Wertebereich ein ausgeprägtes nichtlineares Verhalten besitzen.

Als Aktivierungsfunktionen (AF) werden Sigmoid-Funktionen verwendet, da in [129] gezeigt werden konnte, dass diese im Kontext der dynamischen Reibmodellierung die besten Ergebnisse erwarten lassen. Aktivierungsfunktionen vom Typ der Sigmoid-Funktionen zeichnen sich dadurch aus, dass sie reell, beschränkt, differenzierbar und streng monoton sind. In dem in dieser Arbeit betrachteten KNN wird in der Hiddenschicht der Tangens-Hyperbolicus als AF verwendet:

$$\tanh_v(\alpha) = \frac{2}{1 + e^{-v \cdot \alpha}} - 1 \quad (3.41)$$

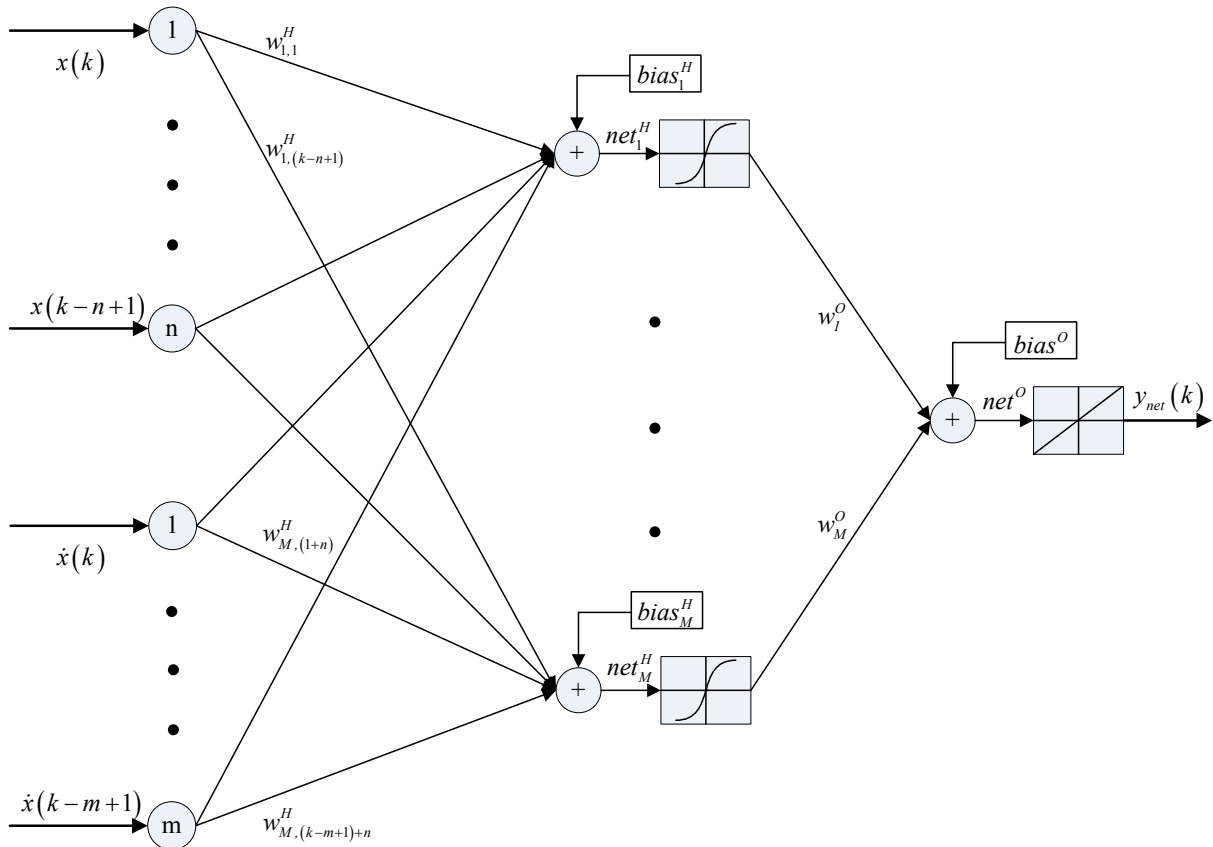


Abbildung 3.15 Struktur des künstlichen neuronalen Netzes

Die Konstante $v \in (0, \infty)$ entspricht dem Anstieg der Sigmoid-Funktion. Für $v \rightarrow \infty$ nähert sich die Sigmoid-Funktion dem normierten Sprung an. Der Parameter v ermöglicht es somit, Einfluss auf die Stärke des nichtlinearen Verhaltens zu nehmen. Da durch das Anlernen der Gewichte dieser Anstieg indirekt vom Netz selbst festgelegt wird, hat der Parameter für das angelernte Netz keinen direkten Einfluss. Daher wird dieser Wert auf $v = 1$ festgelegt. Der Eingang α wird auf einen Wertebereich von $(-1, 1)$ normiert.

Für die Berechnung der Ausgabe des i -ten Neurons in der Hiddenschicht des Hauptnetzes gelten folgende Gleichungen:

$$f_i^H(net_i^H) = \tanh(net_i^H) = \frac{2}{1 + e^{-net_i^H}} - 1 \quad (3.42)$$

$$\text{mit } net_i^H = \sum_{q=1}^n w_{i,q}^H \cdot U_{1,q}(k) + \sum_{p=1}^m w_{i,p+n}^H \cdot U_{2,p}(k) + bias_i^H$$

Hierbei ist net_i^H die gewichtete Summe aller Eingangswerte des KNN addiert mit dem $bias$ -Wert. Der Faktor $w_{i,q}^H$ ist das Gewicht vom q -ten Eingang zum i -ten Neuron der Hiddenschicht. Der $bias$ -Wert stellt den Schwellwert des Hiddenneurons dar und wird während der Trainingsphase mit angepasst. Die Ausgabe des i -ten Neurons der Hiddenschicht im KNN entspricht dem Funktionswert der Aktivierungsfunktion f_i^H .

Die Ausgangswerte der Hiddenneuronen beider Teilnetze werden an das Ausgangsneuron in der Ausgangsschicht weitergeleitet. Dort werden alle Werte gewichtet aufsummiert und über eine lineare AF dem Netzausgang y_{net} übergeben:

$$y_{net} = \sum_{i=1}^M w_i^O \cdot f_i^H(net_i^H) \quad (3.43)$$

w_i^O ist hierbei das Gewicht zwischen der Ausgangsschicht und dem i -ten Hiddenneuron des KNN.

3.4.3 Parameteridentifikation

Da besonders im Nanometerbereich die dynamischen Reibeffekte nicht vernachlässigt werden dürfen, kommen für eine erfolgreiche Nachbildung des Systemverhaltens von NPM-Systemen nur dynamische Reibmodelle in Frage. Demzufolge werden in diesem Abschnitt das LuGre-, das elasto-plastische-, das GMS- und das DNLRX-Modell exemplarisch für die Y-Achse des xy-Demonstrators identifiziert. Weiterhin wird das in Abschnitt 3.4.2.2 vorgestellte künstliche neuronale Netz zur Nachbildung der Reibung eingesetzt. Im Folgenden wird für jedes betrachtete Modell das prinzipielle Vorgehen bei der Parameteridentifikation kurz skizziert. Darüber hinaus wird detailliert auf die eingesetzten Optimierungsverfahren eingegangen, da deren Auswahl und Konditionierung einen entscheidenden Einfluss auf die erzielbare Modellgüte besitzt.

3.4.3.1 Identifikation des LuGre- und des elasto-plastischen Reibmodells

Das elasto-plastische Reibmodell ist eine Erweiterung des LuGre-Modells und somit gestaltet sich der Identifikationsprozess nahezu identisch. Aus diesem Grund soll dieser im Folgenden anhand des komplexeren elasto-plastischen Reibmodells vorgestellt werden. Prinzipiell lässt sich die Parameteridentifikation in zwei Abschnitte unterteilen. Der erste Teil dient der Bestimmung der Stribeckkurve, aus welcher sich die Viskositätskonstante σ_2 und alle Parameter der Funktion $F_{ss}(\dot{x}(t))$ (siehe Gleichung (3.27)) ergeben. Gegenstand des zweiten Teils der Parameteridentifikation ist die Analyse des Reibverhaltens im elastischen und elasto-plastischen Bereich. Daraus werden die Federsteifigkeitskonstante σ_0 , die Dämpfungskonstante σ_1 und die kritische Auslenkung z_{ba} abgeleitet. Wie bereits erwähnt wurde, beginnt der Parameteridentifikationsprozess mit der Ermittlung der Stribeckkurve [20]. Hierzu wird die betrachtete Achse des xy-Demonstrators im geregelten Betrieb mit verschiedenen konstanten Geschwindigkeiten verfahren und die dabei auftretende Reibkraft gemessen. Da die Beschleunigungskräfte bei einer Bewegung mit konstanter Geschwindigkeit gleich Null sind, ist der Betrag der Reibkraft gleich dem Betrag der Antriebskraft:

$$F_R(t) = F_A(t) = k_A \cdot k_V \cdot u_S(t) \quad (3.44)$$

Zur Bestimmung der Stribeckkurve wird am xy-Demonstrator ein Geschwindigkeitsbereich von $\pm 10^{-7}$ m/s bis $\pm 10^{-2}$ m/s untersucht [20]. Für jeden aufzunehmenden Datenpunkt⁹ der Stribeckkurve wird ein gesondertes Experiment durchgeführt. Dabei wird die betrachtete Achse des xy-Demonstrators auf eine vorgegebene konstante Geschwindigkeit beschleunigt und anschließend der vom Regelungssystem aufgebrachte Antriebsstrom über eine Zeitspanne von zwei Sekunden gemittelt. Somit steht gemäß Gleichung (3.44) indirekt ein Messwert für die Reibkraft zur Verfügung. Um darüber hinaus die Ergebnisse statistisch bewerten zu können, werden pro Datenpunkt 400 Versuche durchgeführt und der Mittelwert, die Standardabweichung sowie ein Konfidenzintervall ermittelt. Da diese Daten auch das viskose Reibverhalten mit abbilden, wird zur Modellierung des statischen Reibverhaltens anstatt Gleichung (3.21) eine alternative Form verwendet [20]:

$$F_{Stribeck}(\dot{x}(t)) = F_C + F_S \cdot e^{-|\dot{x}(t)/v_s|^\beta} + \sigma_2 \cdot \dot{x}(t) \quad (3.45)$$

$F_{Stribeck}$ beschreibt das Reibverhalten natürlich nur in einer Bewegungsrichtung und daher wird Gleichung (3.45) für jede Bewegungsrichtung separat identifiziert. Aufgrund der Tatsache, dass sich das System im „Rolling“-Bereich befindet, gilt $\dot{z}(t) = 0$. Somit können aus den experimentellen Daten direkt die Modellparameter F_C , F_S , v_s , β und σ_2 mittels einer statischen Optimierung bestimmt werden. Grundlage für den angewendeten Optimierungsalgorithmus ist im Diskreten ein quadratisches Gütekriterium der Form:

$$J = \sum_{k=1}^{N_{ela-pla}} e^2(k) \quad (3.46)$$

⁹ Ein Datenpunkt ist ein Messwertpaar aus Geschwindigkeit und zugehöriger Reibkraft.

Der Modellfehler $e(k)$ wird aus der Differenz zwischen gemessener Reibkraft $F_R(k)$ und berechneter Reibkraft $\hat{F}_R(k)$ gebildet und so ergibt sich:

$$J = \sum_{k=1}^{N_{\text{ela-pla}}} \left(F_R(k) - \hat{F}_R(k) \right)^2 \quad (3.47)$$

Demzufolge kann ein Optimierungsproblem formuliert werden:

$$\left[\hat{F}_C, \hat{F}_S, \hat{v}_S, \hat{\sigma}_2, \hat{\beta} \right] = \arg \min_{F_C, F_S, v_S, \sigma_2, \beta} J(F_C, F_S, v_S, \sigma_2, \beta) \quad (3.48)$$

Mittels eines erweiterten Quasi-Newton-Verfahrens werden die gesuchten Modellparameter anhand der experimentell ermittelten Kraft-Geschwindigkeit-Messwertpaare bestimmt [29]. Die Ergebnisse der Parameterschätzung sind in Tabelle 3.1 zusammengefasst und der Verlauf der Funktion $F_{\text{Stribeck}}(\dot{x}(t))$ ist exemplarisch für die positive Bewegungsrichtung in Abbildung 3.16 dargestellt. Abbildung 3.17 zeigt die sich ergebende Stribeckkurve mit den zugehörigen Standardabweichungen. Wie bereits in Abschnitt 3.4.1 festgestellt wurde, ist eine Stribeck-Überhöhung (siehe Abbildung 3.11 b)) in Abbildung 3.17 nicht erkennbar. Stattdessen sinkt die Reibkraft bei niedrigen Geschwindigkeiten kontinuierlich, was typisch für kugelgelagerte Führungen ist [2, 32].

Im zweiten Teil der Parameteridentifikation werden die Federkonstante σ_0 , die Dämpfungskonstante σ_1 und die kritische Borstenauslenkung z_{ba} bestimmt. Hierzu wird ein angepasstes Experiment im „Pre-Rolling“-Bereich durchgeführt, da die viskosen Eigenschaften des Schmierstoffs in Abhängigkeit von der relativen Geschwindigkeit bei mikroskopischen Bewegungen vernachlässigt werden können. So vereinfacht sich Gleichung (3.23) zu:

$$F_R(t) = \sigma_0 \cdot z(t) + \sigma_1 \cdot \dot{z}(t) \quad (3.49)$$

Weiterhin müssen die Trägheitskräfte im „Pre-Rolling“-Bereich nicht berücksichtigt werden und somit entspricht auch bei diesem Experiment die applizierte Antriebskraft der Reibkraft (siehe Gleichung (3.44)).

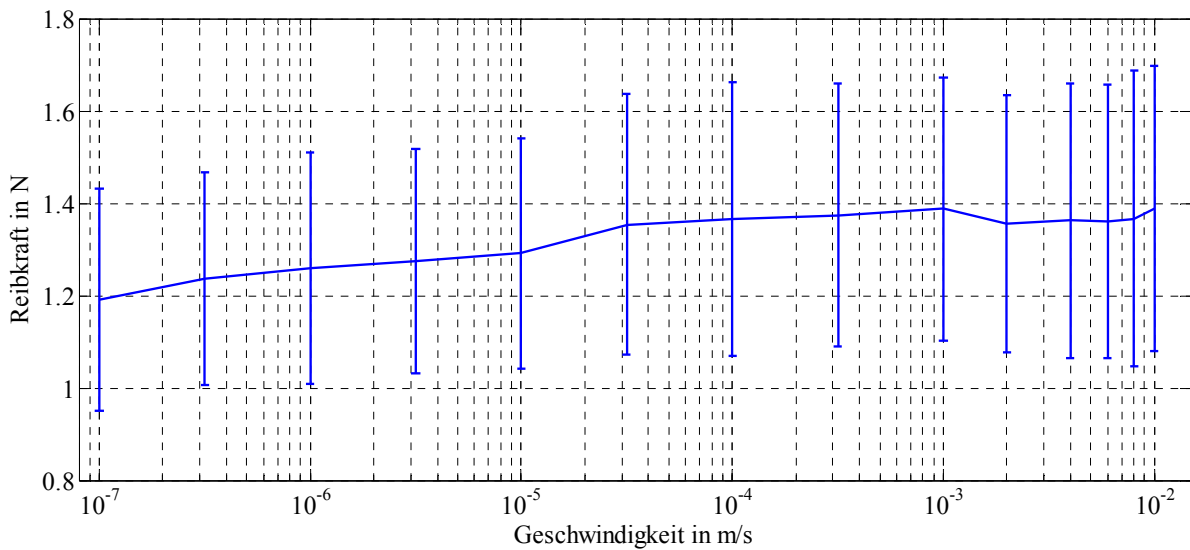


Abbildung 3.16: Approximierte Stribeckkurve für die positive Bewegungsrichtung in logarithmischer Darstellung mit zugehörigen Standardabweichungen

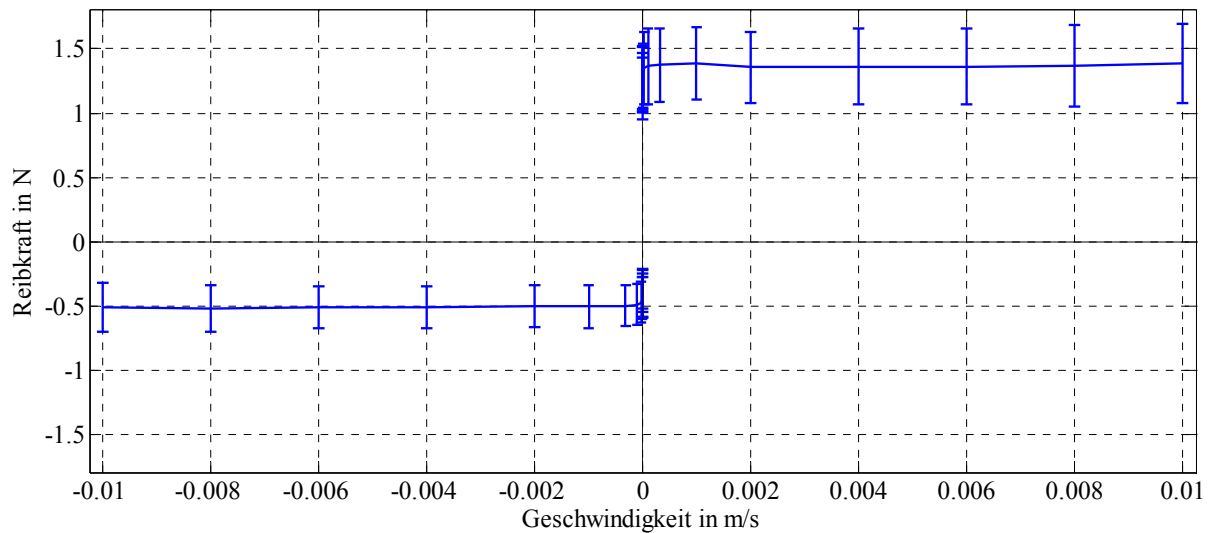


Abbildung 3.17: Ermittelte Stribeckkurve der Y-Achse mit zugehörigen Standardabweichungen

Um die gesuchten Parameter σ_0 , σ_1 und z_{ba} zu bestimmen, werden die Linearmotoren der betrachteten Achse mit trapezförmigen Stromverläufen unterschiedlicher Amplitude beaufschlagt. Die daraus resultierenden Kraft-/Positionsverläufe werden anschließend zur Parameteridentifikation verwendet. Abbildung A 12 und Abbildung A 13 zeigen beispielhaft die Stromtrajektorie sowie den zugehörigen Positionsverlauf bei einer maximalen Antriebskraft von 0,126 N. Wird die Antriebskraft gegen die Auslenkung aufgetragen (siehe Abbildung 3.18), ist deutlich zu erkennen, dass das Verhalten der Reibkraft einer nichtlinearen Feder entspricht. Plastische beziehungsweise irreversible Verschiebungen treten bei diesem Experiment nicht auf.

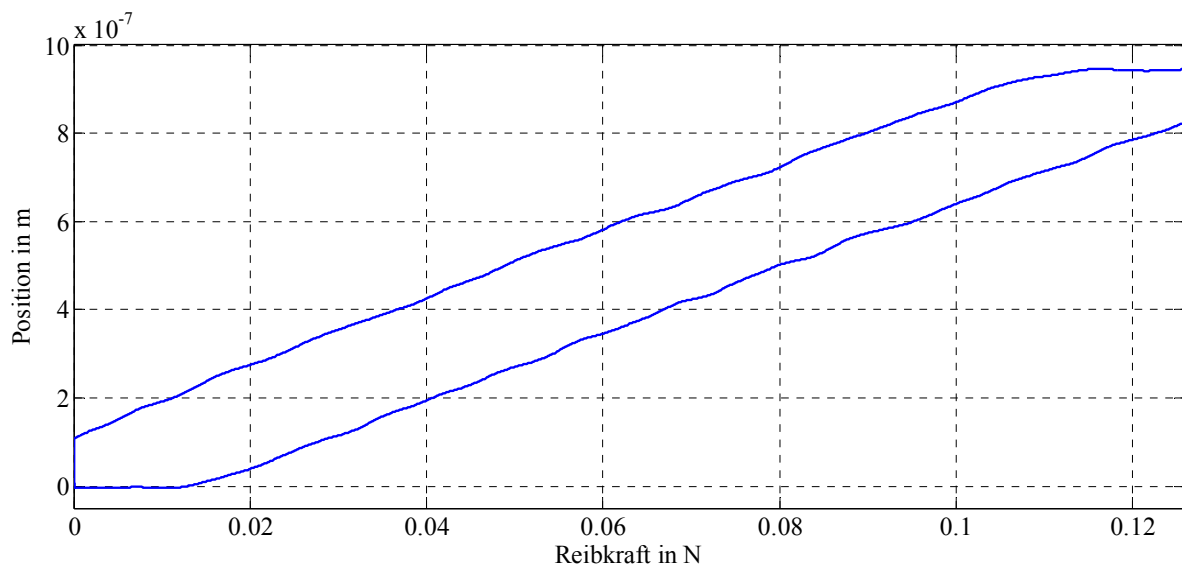


Abbildung 3.18: Reibkraft in Abhängigkeit der Position im „Pre-Rolling“-Bereich zur Bestimmung von σ_1 , σ_2 und z_{ba}

Auch hier liegt der Parameterschätzung das quadratische Gütemaß aus Gleichung (3.46) zu Grunde und es kann unter Nutzung der bereits identifizierten Modellparameter folgendes Optimierungsproblem formuliert werden:

$$[\hat{\sigma}_0, \hat{\sigma}_1, \hat{z}_{ba}] = \arg \min_{\sigma_0, \sigma_1, z_{ba}} J(\sigma_0, \sigma_1, z_{ba}), \quad \text{mit } \sigma_0, \sigma_1, z_{ba} > 0, z_{ba} \cdot \sigma_0 \leq F_{ss}(\dot{x}(t)) \quad (3.50)$$

Das Optimierungsproblem wird ebenfalls mit einem erweiterten Quasi-Newton-Verfahren gelöst [29]. Alle identifizierten Modellparameter sind in Tabelle 3.1 dargestellt.

Der Parameteridentifikationsprozess für das LuGre-Modell gestaltet sich nahezu gleich. Begonnen wird mit der experimentellen Ermittlung der Stribeckkurve. Auch beim LuGre-Modell kommt anstelle von Gleichung (3.21) die Gleichung (3.45) zur Anwendung und es werden die Modellparameter F_C und σ_2 mittels einer statischen Optimierung berechnet. Im Anschluss daran erfolgt die Bestimmung der Federkonstante σ_0 und der Dämpfungskonstante σ_1 anhand des bereits vorgestellten Experimentes im „Pre-Rolling“-Bereich. Somit entsprechen alle Parameter des LuGre-Modells denen des elasto-plastischen Reibmodells, und es sei in diesem Zusammenhang ebenfalls auf Tabelle 3.1 verwiesen.

Während des Identifikationsprozesses fiel auf, dass die Parameter sehr stark streuen. Dies ist ein typisches Merkmal von kugelgeführten NPM-Systemen und muss in einem zu entwerfenden modellbasierten Regelungssystem natürlich berücksichtigt werden (siehe Kapitel 4).

Tabelle 3.1: Parameter für das LuGre- und das elasto-plastische Reibmodell

Parameter	Wert
F_{C+}	1,371 N \pm 0,276 N
F_{C-}	-0,515 N \pm 0,165 N
F_{S+}	-0,211 N
F_{S-}	0,265 N
v_{S+}	$5,22 \cdot 10^{-6}$ m/s
v_{S-}	$1,353 \cdot 10^{-6}$ m/s
σ_{2+}	0,141 Ns/m
σ_{2-}	0,298 Ns/m
β_+	0,371
β_-	0,194
σ_0	1.217.400 N/m
σ_1	160.094 Ns/m
z_{ba}	$2,001 \cdot 10^{-6}$ m

3.4.3.2 Identifikation des GMS-Modells

Identifikationsalgorithmus

Da sowohl der „Maxwell-Slip“-Ansatz als auch die Borstenmodelle auf der Stribeckkurve basieren, beginnt der Parameteridentifikationsprozess des GMS-Modells ebenfalls mit der Ermittlung dieser statischen Kennlinie. Das Vorgehen entspricht dem bereits im vorangegangenen Abschnitt beschriebenen Algorithmus und soll daher an dieser Stelle nicht noch einmal explizit erläutert werden. Daher wurde die bereits identifizierte Stribeckkurve aus Abbildung 3.17 auch für das GMS-Modell verwendet. In der zweiten Phase der Parameteridentifikation müssen die verbleibenden Parameter bestimmt werden. Hierbei handelt es sich um den Attraktionsparameter C , die Skalierungsfaktoren γ_i , die Dämpfungskonstanten σ_i und die Federsteifigkeiten k_i sowie die Steigung des viskosen Reibanteils k_v . Problematisch bei der Identifikation dieser Parameter ist die Tatsache, dass der Attraktionsparameter C und die Skalierungsfaktoren γ_i nichtlinear in das Modell eingehen und kein Experiment gefunden werden kann, welches es erlaubt, diese Parameter unabhängig von den verbleibenden linearen Parametern zu identifizieren. Somit müssen alle Modellparameter gemeinsam aus einem Identifikationsdatensatz heraus ermittelt werden. Zur Lösung des Problems wird eine Strategie eingesetzt, welche von Rizos *et al.* im Zusammenhang mit der Identifikation des DNLRX-Reibmodells vorgeschlagen wurde [6, 103, 104, 142]. Wie bereits bei der Identifikation der Stribeckkurve basiert der eingesetzte Identifikationsalgorithmus auf einem quadratischen Gütekriterium:

$$J = \sum_{k=1}^{N_{GMS}} e^2(k) \quad (3.51)$$

Gemäß Gleichung (3.1) lässt sich die Reibkraft $F_R(k)$ in diskreter Zeit berechnen aus:

$$F_R(k) = F_A(k) - m \cdot a(k) \quad (3.52)$$

Der Modellfehler $e(k)$ wird aus der Differenz zwischen rekonstruierter Reibkraft $F_R(k)$ und berechneter Reibkraft $\hat{F}_R(k)$ gebildet und so gilt:

$$J = \sum_{k=1}^{N_{GMS}} \left(F_R(k) - \hat{F}_R(k) \right)^2 \quad (3.53)$$

Der Attraktionsparameter C und die Skalierungsfaktoren γ_i werden in einem $(M+1)$ -dimensionalen Parametervektor $\mathbf{T} = [\gamma_1 \dots \gamma_M : C]$, der nichtlinear in $\hat{F}_R(k)$ eingeht, hinterlegt. Werden darüber hinaus alle verbleibenden Parameter in einem $(2M+1)$ -dimensionalen Vektor $\Phi = [\sigma_1 \dots \sigma_M : k_1 \dots k_M : k_v : b]^T$ zusammengefasst, ist es möglich, ein sequentielles zweistufiges Optimierungsproblem zu formulieren:

$$[\hat{\Phi} \quad \hat{\mathbf{T}}] = \arg \min_{\mathbf{T}} \left\{ \min_{\Phi} J(\Phi, \mathbf{T}) \right\} \quad (3.54)$$

Der Parametervektor Φ beinhaltet zusätzlich zu den linearen Modellparametern einen linearen Modelloffset b , welcher der Abbildung von systematischen Fehlern dient¹⁰. Basierend auf Gleichung (3.54) können die Modellparameter anhand von Paaren aus Kraft- und zugehörigen Positionsmesswerten bestimmt werden. Im ersten Schritt der Parametersuche lokalisiert ein evolutionäres Suchverfahren im Parameterraum Bereiche mit lokalen Minima [53]. Anschließend wird ein Regressionsalgorithmus genutzt, um in den zuvor bestimmten Regionen das globale Minimum zu finden [132]. Mit Bezug auf das hier eingesetzte GMS-Modell wird durch den evolutionären Algorithmus der nichtlineare Parametervektor T bestimmt. Daran anknüpfend identifiziert der Regressionsalgorithmus die verbleibenden linearen Modellparameter im Parametervektor Φ . Dieser alternierende Prozess wird so lange wiederholt, bis eine a priori definierte Fehlerschranke unterschritten wird.

Problematisch bei der Parameteridentifizierung sind die Anfangswerte der Federauslenkungen δ_i . Aufgrund der Tatsache, dass diese rekursiv berechnet werden, sind die Anfangswerte $\delta_i(k=1)$ zwingend erforderlich. Eine eindeutige Bestimmung der Startwerte durch Parameterschätzung ist nicht möglich, und eine zufällige Initialisierung führt zu einer Verschlechterung der erreichbaren Modellgüte. Um dies zu vermeiden, kann eine besondere Eigenschaft der „Maxwell-Slip“-Modelle ausgenutzt werden. Wird aus dem Identifizierungsdatensatz ein Kraft-/Positions-Messwertpaar ausgewählt, an dem sich das System definitiv im „Rolling“-Bereich befindet, gilt zu diesem definierten Zeitpunkt t_{rol} im Falle des GMS-Modells:

$$\delta_i = \text{sgn}[x(t_{rol})] \cdot \gamma_i \cdot g(\dot{x}(t)) \quad (3.55)$$

In der praktischen Umsetzung wird im Identifikationsdatensatz der Zeitpunkt des globalen Positionsextremwertes zu t_{rol} gewählt, da sich an diesem Punkt das System eindeutig im „Rolling“-Bereich befindet. Nun kann Gleichung (3.61) Anwendung finden und im folgenden Identifikationsprozess werden nur Datensätze mit $t > t_{rol}$ verwendet.

Die Modellordnung des GMS-Modells wird iterativ während des Identifikationsprozesses festgelegt. Hierbei wird die beschriebene Parameteridentifikation mehrfach durchgeführt und dabei die Modellordnung schrittweise verringert. Erst wenn sich die Güte des Modells signifikant verschlechtert, wird der Prozess gestoppt. Um die Qualität des Parameteridentifikationsprozesses zu bewerten, wird der Root Mean Square Error (RMSE) verwendet. Dieser ist definiert als:

$$\text{RMSE} = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N (F_R(k) - \hat{F}_R(k))^2} \quad (3.56)$$

N ist hierbei die Anzahl der Wertepaare im Identifikationsdatensatz, $F_R(k)$ die gemessene und $\hat{F}_R(k)$ die geschätzte Kraft.

¹⁰ Verkippungen des experimentellen Aufbaues können beispielsweise systematische Fehler verursachen.

Identifikationsergebnisse

Für die Parameteridentifikation des GMS-Modells wurden verschiedene Trajektorien im gesteuerten Betrieb aufgenommen, um das Systemverhalten ohne Regelungssystem abbilden zu können. Abbildung 3.19 zeigt den Identifikationsdatensatz, der sich aus N Wertepaaren, bestehend aus erregender Kraft $F_A(k)$ (siehe Abbildung 3.19 a) und resultierender Position $x(k)$ (siehe Abbildung 3.19 b), zusammensetzt. Das Kraftsignal in Abbildung 3.19 ist aufgebaut aus einer Sinuskurve ($f=1\text{Hz}$) und einem tiefpassgefilterten Rauschen. Das Signal ist so gewählt worden, dass das System gleich zu Beginn sehr stark ausgelenkt wird. Damit befindet sich der Positioniertisch im „Rolling“-Bereich, was für die Bestimmung von t_{rol} entscheidend ist. Der dargestellte Signalverlauf stellt natürlich nur ein Beispiel für eine mögliche Reaktion des Systems dar, denn die Startposition hat einen signifikanten Einfluss auf den Verlauf des Positionssignals. Der Extremwert des Positionssignals liegt bei $3,873 \cdot 10^{-6}\text{ m}$ und der daraus resultierende Zeitpunkt t_{rol} wurde in Abbildung 3.19 b) kenntlich gemacht. Auf Basis dieses Identifikationsdatensatzes wurde die Reibkraft $F_R(k)$ mithilfe von Gleichung (3.52) rekonstruiert. Während des beschriebenen Identifikationsprozesses wurden die besten Ergebnisse mit sieben „Maxwell-Slip“-Elementen erreicht. Das Potential des GMS-Modells zur Nachbildung des Systemverhaltens ist in Abbildung 3.20 dargestellt. Der erreichte RMSE liegt hier bei $0,091\text{ N}$.

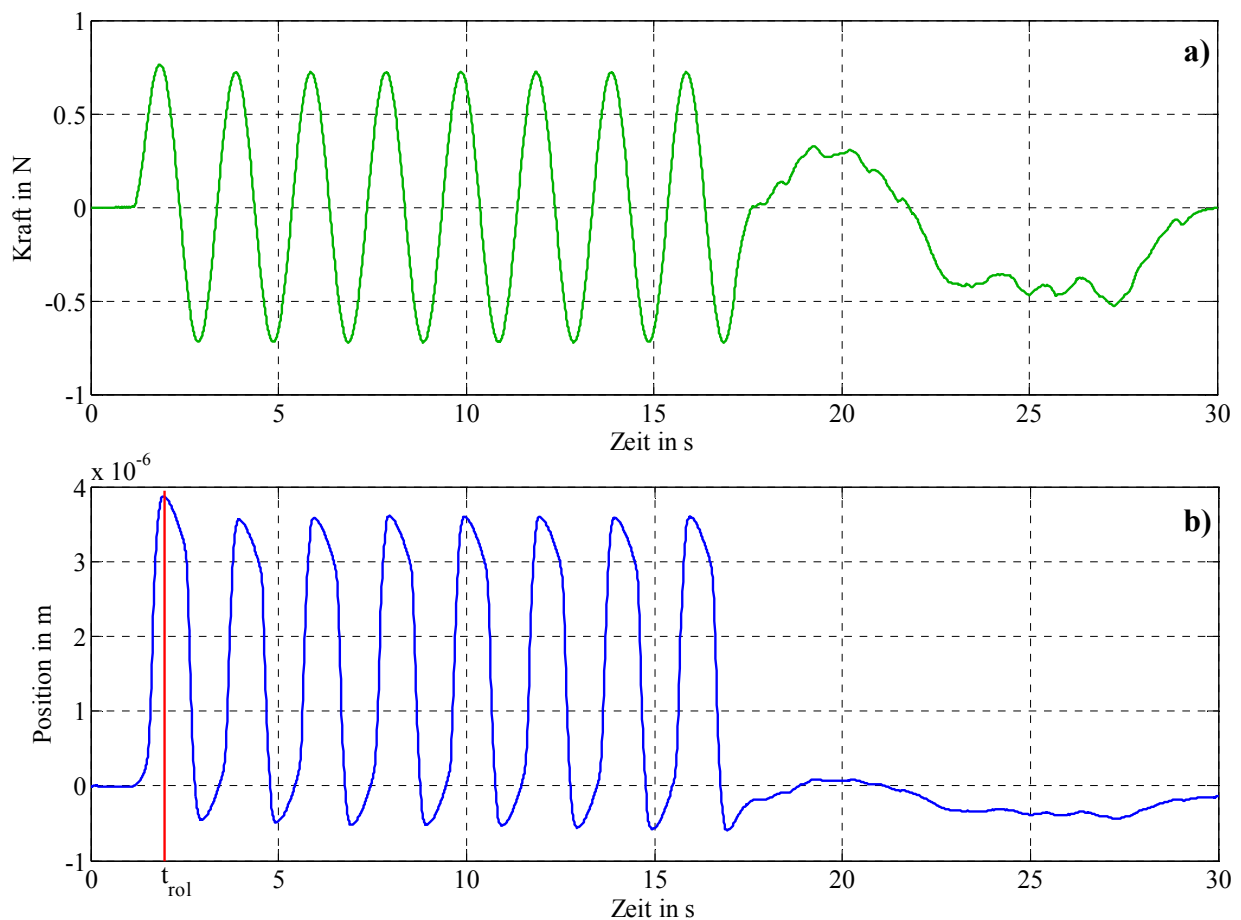


Abbildung 3.19: Identifikationsdatensatz: (a) erregende Kraft; (b) gemessene Position

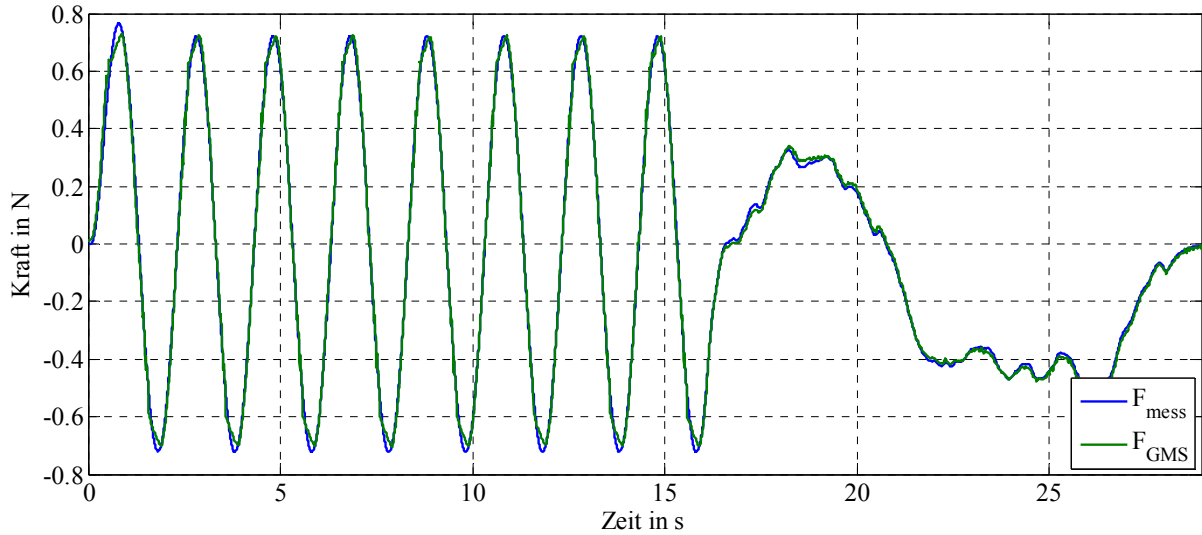


Abbildung 3.20: Identifikationsdatensatz: Vergleich zwischen gemessener und geschätzter Kraft

3.4.3.3 Identifikation des DNLRX-Modells

Identifikationsalgorithmus

Analog zum GMS-Modell ist der ausgesprochen große Parameterumfang des DNLRX-Reibmodells das primäre Problem, was es im Zuge der Parameteridentifikation zu lösen gilt. Auch hier müssen sowohl lineare als auch nichtlineare Modellparameter gemeinsam aus einem Identifikationsdatensatz heraus ermittelt werden. Dem zweistufigen Identifikationsalgorithmus aus Abschnitt 3.4.3.2 liegt ebenfalls ein quadratisches Gütekriterium zugrunde [6, 103, 104, 142]:

$$J = \sum_{k=1}^{N_{DNLRX}} e^2(k) \quad (3.57)$$

Da das DNLRX-Modell neben der Reibung auch die Trägheit des kinematischen Systems abbildet, kann gemäß Gleichung (3.1) der Modellfehler $e(k)$ aus der Differenz zwischen gemessener Antriebskraft $F_A(k)$ und berechneter Antriebskraft $\hat{F}_A(k)$ gebildet werden:

$$J = \sum_{k=1}^{N_{DNLRX}} \left(F_A(k) - \hat{F}_A(k) \right)^2 \quad (3.58)$$

Wird die Berechnungsvorschrift des DNLRX-Modells (siehe Gleichung (3.39)) in Gleichung (3.58) eingesetzt, so ergibt sich:

$$\text{DNLRX}(M, n_\delta, n_x, \mathbf{\Delta}, \mathbf{\Theta}) : F_A(k) = \mathbf{\Phi}^T \cdot \left[x(k) \dots x(k-n_x) ; \delta^T(k) \dots \delta^T(k-n_\delta) ; 1 \right]^T + e(k) \quad (3.59)$$

Der Parametervektor der Dimension $[(n_\delta+1) \cdot M + n_x + 2]$ beinhaltet alle zu identifizierenden Modellparameter sowie einen zusätzlichen Modelloffset b . Der Parametervektor $\mathbf{\Phi}$ geht linear und der M -dimensionale Schwellwertvektor, $\mathbf{\Delta} = [\Delta_1 \dots \Delta_M]$ nichtlinear in die Modellgleichungen ein. Somit kann wie beim GMS-Modell ein sequentielles zweistufiges Optimierungsproblem formuliert werden:

$$\begin{bmatrix} \hat{\Phi} & \hat{\Delta} \end{bmatrix} = \arg \min_{\Delta} \left\{ \min_{\Phi} J(\Phi, \Delta) \right\} \quad (3.60)$$

Nun kommt das schon im vorangegangenen Abschnitt beschriebene hybride Parametersuchverfahren zum Einsatz. Auch beim DNLRX-Modell stellt sich die Frage nach geeigneten Startwerten für die Federauslenkungen δ_i . Deren Bestimmung erfolgt analog zum GMS-Modell, wobei sich Gleichung (3.55) vereinfacht zu:

$$\delta_i = \text{sgn}[x(t_{rol})] \cdot A_i \quad (3.61)$$

Identifikationsergebnisse

Für die Parameteridentifikation des DNLRX-Modells wurde wie im vorangegangenen Abschnitt der Identifikationsdatensatz aus Abbildung 3.19 verwendet. Auf Basis dieses Datensatzes wurden während des Identifikationsprozesses die besten Ergebnisse mit einer minimalen Modellordnung von $M = 8$, $n_x = 5$, $n_\delta = 0$ erzielt [5, 141]. Das Potential des DNLRX-Modells zur Nachbildung des Systemverhaltens ist in Abbildung 3.21 dargestellt. Der RMSE liegt hier bei 0,008 N.

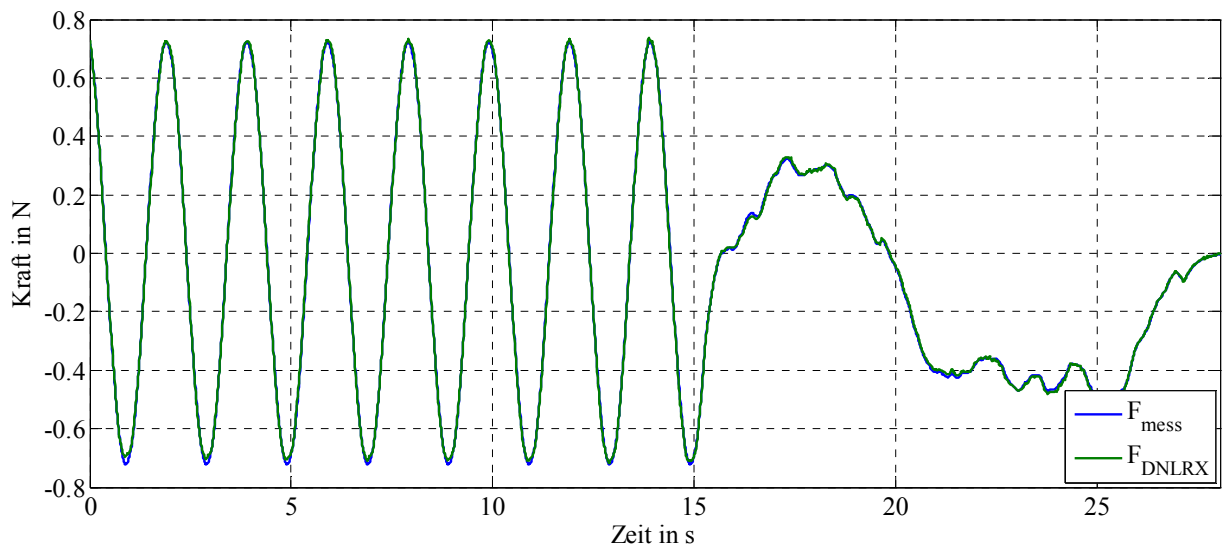


Abbildung 3.21: Identifikationsdatensatz: Vergleich zwischen gemessener und geschätzter Kraft

3.4.3.4 Identifikation des künstlichen Neuronen Netzes

Identifikationsalgorithmus

Ein künstliches Neuronales Netz besitzt prinzipbedingt keinerlei a priori Information über das Verhalten, welches es reproduzieren soll. Bevor also ein KNN eine geforderte Aufgabe erfüllen kann, muss es angelernet werden. Für das in Abschnitt 3.4.2.2 vorgestellte KNN wird als Lernverfahren der Levenberg-Marquardt-Algorithmus vorgeschlagen [80, 129, 134]. Dieser Optimierungsalgorithmus gehört zu der Klasse der zielorientierten Lernverfahren und zeichnet sich durch eine hohe Konvergenz aus. Der Algorithmus kombiniert das Gauß-Newton-Verfahren mit einer Regularisierungstechnik, die die Robustheit des Verfahrens deutlich steigert. Weiterhin erreicht der Levenberg-Marquardt-Algorithmus eine wesentlich höhere Lern-

geschwindigkeit als der klassische Back Propagation Algorithmus (BPA) [78]. Dies wird erreicht, indem höhere Ableitungen der Fehlerfunktion in die Berechnung der Gewichtsänderungen eingehen. Nachteilig wirkt sich jedoch der verhältnismäßig hohe Rechenaufwand aus, da eine Inversion der Jacobi-Matrix durchgeführt werden muss.

Prinzipiell lässt sich der Levenberg-Marquardt-Algorithmus in drei Berechnungsschritte unterteilen. Im ersten Schritt wird die Netzausgabe für einen oder mehrere¹¹ Eingangsdatensätze berechnet. Daraufhin wird im zweiten Schritt auf Basis des Ausgangsfehlers die Änderung aller Gewichte im KNN bestimmt. Im letzten Berechnungsschritt wird der Lernerfolg ermittelt und in dessen Abhängigkeit die Schrittweite für die nächste Gewichts Anpassung berechnet. Der Algorithmus endet, wenn ein bestimmtes Zielkriterium erfüllt ist. Im Folgenden werden die mathematischen Grundlagen des Levenberg-Marquardt-Algorithmus kurz skizziert. Der Ausgangspunkt des Verfahrens ist der Gradientenabstieg in einem Gütegebirge. Die Basis bildet hierbei folgende globale Gütefunktion:

$$J(\mathbf{w}) = \sum_{k=1}^N \left[\sum_{p=1}^M E_p(k) \right] \quad (3.62)$$

Der Parameter N entspricht der Anzahl der Identifikationsdatensätze einer Lernepoche und M der Menge der Ausgangsneuronen. Der Gewichtsvektor $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_o]^T$ besteht aus allen Gewichten des Netzes, die durch den Lernalgorithmus angepasst werden. Als Fehlerfunktion $E_p(k)$ wird der mittlere quadratische Fehler zwischen Netzausgang und gewünschtem Ausgang verwendet. Hierbei ist $y_p(k)$ der gewünschte Netzausgang für das k -te Muster am p -ten Ausgangsneuron:

$$E_p(k) = (y_p(k) - y_{net,p}(k))^2 \quad (3.63)$$

Da das in dieser Arbeit betrachtete neuronale Netzwerk lediglich ein Ausgangsneuron besitzt, gilt $p = 1$ und es soll im Folgenden gelten:

$$E_1(k) = (y_1(k) - y_{net,1}(k))^2 = E(k) = (y(k) - y_{net}(k))^2 \quad (3.64)$$

Der zu berechnende Gradient ist die Ableitung der Fehlerfunktion $E(k)$ nach allen Gewichten. In der folgenden Herleitung ist f^O die Aktivierungsfunktion der Ausgangsschicht, f^H die Aktivierungsfunktion der Hiddenschicht, net^H die gewichtete Summe aller Netzeingangswerte in der Hiddenschicht, net_i^H die gewichtete Summe aller Netzeingangswerte für das i -te Neuron der Hiddenschicht und y_{net} die Ausgabe des Netzes. Für die Ausgangsschicht ergibt sich somit:

$$\frac{\partial E(k)}{\partial w_i^O} = \frac{\partial E(k)}{\partial net^O} \cdot f^H(net_i^H) \quad (3.65)$$

¹¹ Das Lernen auf Basis mehrerer Musterdatensätze wird als epochenweises Lernen bezeichnet.

$$\begin{aligned}
\frac{\partial E(k)}{\partial net^O} &= \frac{\partial}{\partial net^O} \cdot (y - y_{net})^2 \\
&= \frac{\partial}{\partial net^O} \cdot (y - f^O(net^O))^2 \\
&= -2 \cdot (y - y_{net}) \cdot f^{O'}(net^O)
\end{aligned} \tag{3.66}$$

Unter Verwendung der Identität $f^O(net^O)$ als Aktivierungsfunktion in der Ausgangsschicht gilt $f^{O'}(net^O) = 1$. Der Gradient für das i -te Gewicht der Ausgangsschicht kann daher wie folgt bestimmt werden:

$$\begin{aligned}
w_{\Delta}^O &= \frac{\partial E(k)}{\partial w_i^O} \\
&= -2 \cdot (y - y_{net}) \cdot f^H(net_i^H)
\end{aligned} \tag{3.67}$$

Zur Berechnung des Gradienten in der Hiddenschicht wird der Fehler am Ausgang dieser Schicht benötigt. Der Fehler steht jedoch nicht direkt zur Verfügung und muss daher vom Netzausgang in Richtung des Netzeinganges zurückpropagiert werden. Mit Zurückpropagieren wird in diesem Zusammenhang ein Mechanismus bezeichnet, der eine gewichtete Aufteilung des Ausgangsfehlers eines Neurons auf die Neuronen der davorliegenden Schicht ermöglicht. Um den Gradienten zu berechnen, kann die folgende Beziehung verwendet werden:

$$\begin{aligned}
\frac{\partial E(k)}{\partial w_{i,q}^H} &= \frac{\partial E(k)}{\partial net_i^H} \cdot \frac{\partial net_i^H}{\partial w_{i,q}^H} \\
&= \frac{\partial E(k)}{\partial net_i^H} \cdot U_q(k)
\end{aligned} \tag{3.68}$$

Anschließend wird $\frac{\partial E}{\partial net_i^H}$ mit ∂f^H erweitert und es gilt:

$$\begin{aligned}
\frac{\partial E(k)}{\partial net_i^H} &= \frac{\partial E(k)}{\partial f^H} \cdot \frac{\partial f^H}{\partial net_i^H} \\
&= \frac{\partial E(k)}{\partial f^H} \cdot \partial f^{H'}(net_i^H)
\end{aligned} \tag{3.69}$$

Für $\frac{\partial E(k)}{\partial f^H}$ ergibt sich somit:

$$\begin{aligned}
\frac{\partial E(k)}{\partial f^H} &= \frac{\partial}{\partial f^H} \cdot (y - f^O(net^O))^2 \\
&= \frac{\partial}{\partial f^H} \cdot \left(y - f^O \left(\sum_{i=1}^N (w_i^O \cdot f^H(net_i^H)) \right) \right)^2 \\
&= -2 \cdot (y - y_{net}) \cdot \sum_{i=1}^N (w_i^O)
\end{aligned} \tag{3.70}$$

Nach Einsetzen von Gleichung (3.70) in Gleichung (3.69) folgt schließlich für $\frac{\partial E(k)}{\partial net^H}$:

$$\frac{\partial E(k)}{\partial net^H} = -2 \cdot (y - y_{net}) \cdot \sum_{i=1}^N w_i^O \cdot f^{H'}(net_i^H) \quad (3.71)$$

Damit kann der Gradient eines Gewichtes eines Neurons der Hiddenschicht mithilfe von Gleichung (3.68) berechnet werden:

$$\begin{aligned} w_{\Delta}^H &= \frac{\partial E(k)}{\partial w_{i,q}^H} \\ &= -2 \cdot (y - y_{net}) \cdot \sum_{i=1}^N w_i^O \cdot f^{H'}(net_i^H) \cdot U(k) \end{aligned} \quad (3.72)$$

Mit Gleichung (3.67) und (3.72) stehen nun Berechnungsvorschriften für die Gradienten aller Gewichte im KNN zur Verfügung. Somit kann nun die Jacobi-Matrix für alle $O = M \cdot (1+n+m)$ Gewichte des KNN für die betrachtete Lernepoche aufgestellt werden:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial E(k=1)}{\partial w_1} & \frac{\partial E(k=1)}{\partial w_2} & \dots & \frac{\partial E(k=1)}{\partial w_O} \\ \frac{\partial E(k=2)}{\partial w_1} & \frac{\partial E(k=2)}{\partial w_2} & \dots & \frac{\partial E(k=2)}{\partial w_O} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial E(k=N)}{\partial w_1} & \frac{\partial E(k=N)}{\partial w_2} & \dots & \frac{\partial E(k=N)}{\partial w_O} \end{bmatrix} \quad (3.73)$$

Basierend auf dieser Jacobi-Matrix kann eine Vorschrift zur Anpassung der Gewichte \mathbf{w} formuliert werden:

$$\mathbf{w}_{e+1} = \mathbf{w}_e - (\mathbf{J}_e^T \cdot \mathbf{J}_e + l_r \cdot \mathbf{I})^{-1} \cdot \mathbf{J}_e^T \cdot \mathbf{E}_e \quad (3.74)$$

Hierbei ist l_r die Lernrate und $\mathbf{E}_e = [E(1) \ E(2) \ \dots \ E(N)]$ beinhaltet den quadratischen Ausgangsfehler für jeden Tastschritt in der betrachteten Epoche. Die Lernrate l_r kann statisch oder dynamisch gewählt werden und repräsentiert die Schrittweite bei der Parametersuche. Nachdem alle Gewichtsänderungen berechnet wurden, wird der Fehler zwischen Netzausgabe und der gewünschten Ausgabe berechnet. Wird der Ausgangsfehler kleiner, war das Lernen erfolgreich und die alten Gewichte werden mit den neu berechneten Gewichten überschrieben.

Identifikationsergebnisse

Das in Abschnitt 3.4.2.2 vorgestellte künstliche Neuronale Netz wird analog zum DNLRX- und GMS-Modell mit dem Datensatz aus Abbildung 3.19 angelernt. Als Eingangsdaten dienen dem Netz die aktuelle und die vergangene Position sowie die dazugehörigen Geschwindigkeitswerte. Die Ausgangsgröße des Netzes ist die im betrachteten Tastschritt wirkende Antriebskraft. Der Identifikationsdatensatz wird normiert und anschließend in einen Muster- sowie in einen Testdatensatz geteilt. Nur der Musterdatensatz wird zum Lernen verwendet

und der Testdatensatz dient zur Überprüfung des Lernerfolges, indem in jedem Identifizierungsdurchlauf die Generalisierungsfähigkeit des KNN anhand des Testdatensatzes überprüft wird. Durch diese Vorgehensweise wird ein Overfitting des KNN vermieden, denn sobald der Validierungsfehler nicht weiter sinkt, wird das Training abgebrochen und das Netz als vollständig angelernt angesehen.

Der beschriebene Identifikationsprozess wurde mehrmals durchgeführt und die Anzahl der Neuronen sowie die Lernrate l_r variiert. Zur Gütebewertung wurde auch hier der RMSE aus Gleichung (3.56) verwendet. Es zeigte sich, dass eine minimale Netzgröße von $N=M=10$ Neuronen zu den besten Resultaten führt. Des Weiteren ergab sich eine optimale Lernrate von $l_r = 0,001$ [129]. Mit dieser Parametrierung gelingt es, den Identifikationsdatensatz mit einem RMSE von 0,0331 N nachzubilden (siehe Abbildung 3.22).

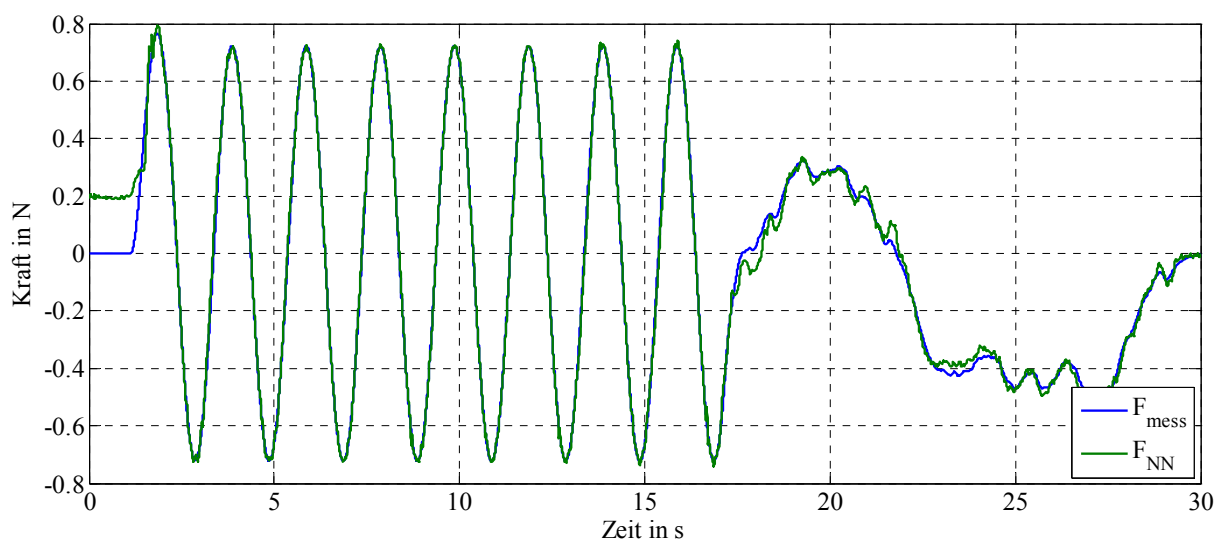


Abbildung 3.22: Identifikationsdatensatz: Vergleich zwischen erregender Kraft und geschätzter Kraft

3.4.4 Leistungsfähigkeit der identifizierten Reibmodelle

In den vorangegangenen Abschnitten konnte gezeigt werden, dass Reibung einen dominanten Einfluss auf das Systemverhalten von kugelgeführten NPM-Systemen ausübt. Basierend auf dieser Erkenntnis wurde das Reibverhalten eingehend phänomenologisch untersucht und festgestellt, dass Reibung ein dynamischer Prozess ist, der besonders im Nanometerbereich ein stark nichtlineares Verhalten aufweist. Daran anknüpfend wurde in Abschnitt 3.4.2 der Stand der Technik auf dem Gebiet der regelungstechnisch angepassten Reibmodellierung dargelegt. Für eine Reibmodellierung im Nanometerbereich kommen nur dynamische Modelle in Frage und daher wurden fünf Modellansätze aus den vergangenen 20 Jahren ausgewählt. Die Identifizierung der Modellparameter wurde in Abschnitt 3.4.3 durchgeführt und im Folgenden sollen nun die identifizierten Modelle validiert werden. Im Anschluss an die Validierung wird die Leistung jedes Modells bewertet und es werden zwei Reibmodelle für eine spätere regelungstechnische Verwendung ausgewählt. Der in Abbildung 3.23 dargestellte Datensatz wur-

de an einer anderen Position¹² als der Identifizierungsdatensatz (siehe Abbildung 3.19) aufgenommen und besteht aus dem Anregungssignal sowie der zugehörigen Systemreaktion.

Auch hier wird zu Beginn das System durch eine sinusförmige Kraft so stark stimuliert, dass es zu plastischen Bewegungen kommt. Gefolgt wird diese Sequenz von einem tiefpassgefilterten Rauschsignal unterhalb der Losbrechkraft. Im Folgenden werden alle betrachteten Reibmodelle anhand des vorgestellten Datensatzes validiert und die Ergebnisse kurz diskutiert. Darauf aufbauend erfolgt in Abschnitt 3.4.4.2 eine Bewertung der Leistungsfähigkeit im Hinblick auf eine möglichst präzise Reibkraftschätzung auf Nanometerebene.

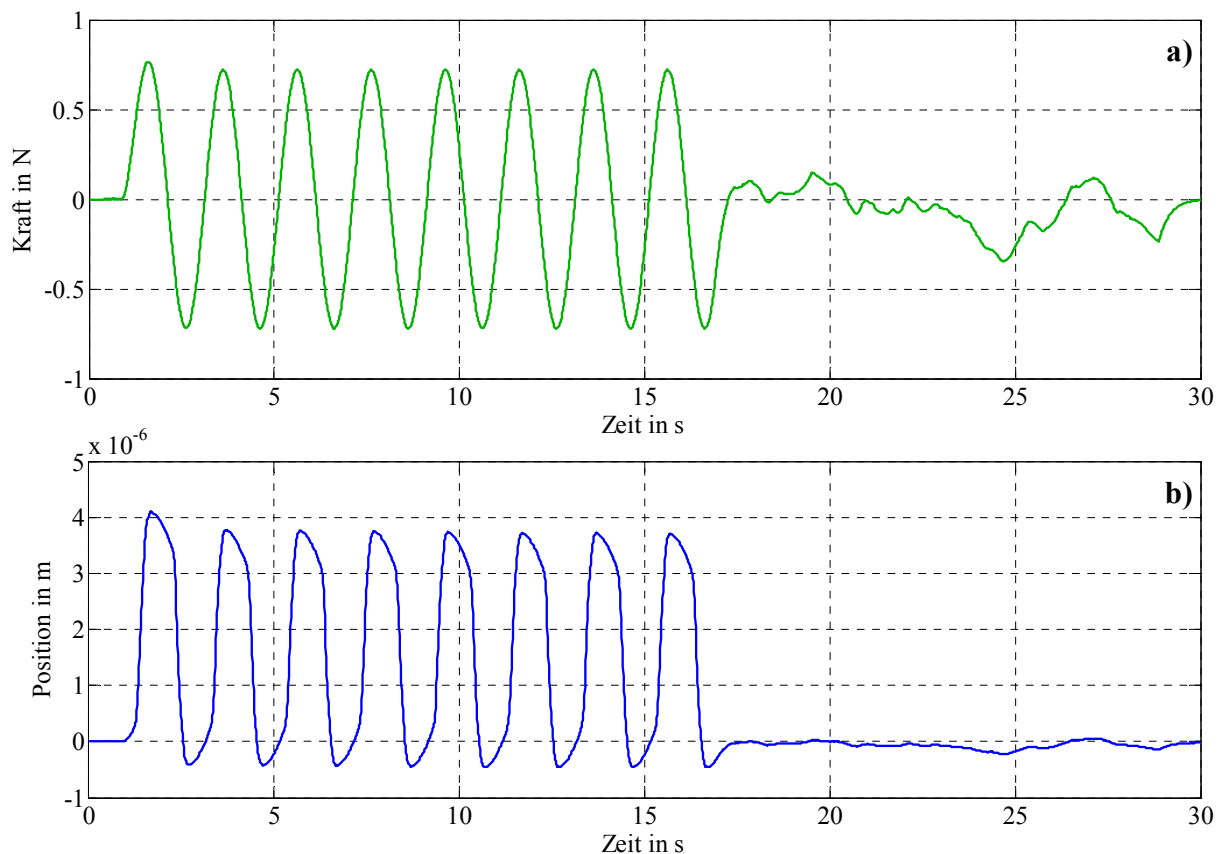


Abbildung 3.23: Validierungsdatensatz: (a) erregende Kraft; (b) gemessene Position

3.4.4.1 Ergebnisse der Validierung

LuGre-Reibmodell

Das LuGre-Modell ist ein reines Reibmodell und so muss zu Beginn der Validierung gemäß Gleichung (3.1) die Trägheitskraft aus dem Validierungsdatensatz heraus gerechnet werden. Im Anschluss daran wird die Geschwindigkeit aus der gemessenen Position rekonstruiert, indem das Positionssignal phasenverzugsfrei gefiltert und anschließend differenziert wird. Der so berechnete Geschwindigkeitsverlauf dient der Anregung des identifizierten LuGre-Modells. Abbildung 3.24 zeigt den Vergleich zwischen gemessener und geschätzter Reibkraft.

¹² Circa 30 mm entfernt von dem Ort an dem der Identifikationsdatensatz aufgenommen wurde.

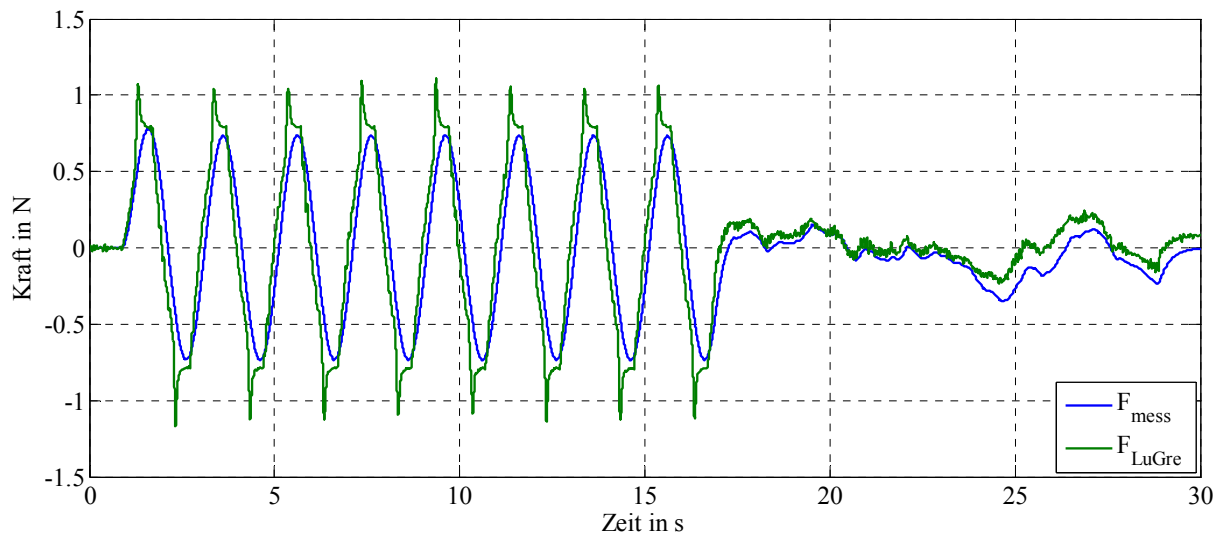


Abbildung 3.24: Vergleich zwischen erregender Kraft und geschätzter Kraft

Anhand der Daten kann festgestellt werden, dass das LuGre-Modell deutliche Schwächen bei der Reibkraftvorhersage nahe der Nullgeschwindigkeit zeigt und somit einen nicht zu vernachlässigenden Modellfehler erzeugt. Besonders in den Umkehrpunkten der Bewegung treten große sprungförmige Fehler auf, die den RMSE bis auf 0,202 N anwachsen lassen. Der Grund für die bedingt zufriedenstellenden Ergebnisse ist das stark mit der Position und Zeit variierende Systemverhalten. Durch diese Schwankungen verlieren die identifizierten Parameter zum Teil ihre Gültigkeit und die Vorhersagegüte des Modells sinkt deutlich.

Elasto-plastisches-Reibmodell

Da das elasto-plastische Reibmodell lediglich eine Erweiterung des LuGre-Modells darstellt, ist die Vorgehensweise bei der Validierung nahezu identisch. Nach der Bereinigung des Validierungsdatensatzes um die Trägheitskraft wird auch hier die Geschwindigkeit genutzt, um das Modell anzuregen. Wieder treten sprungförmige Fehler bei der Richtungsumkehr auf und so wird lediglich ein RMSE von 0,181 N erreicht (siehe Abbildung 3.25).

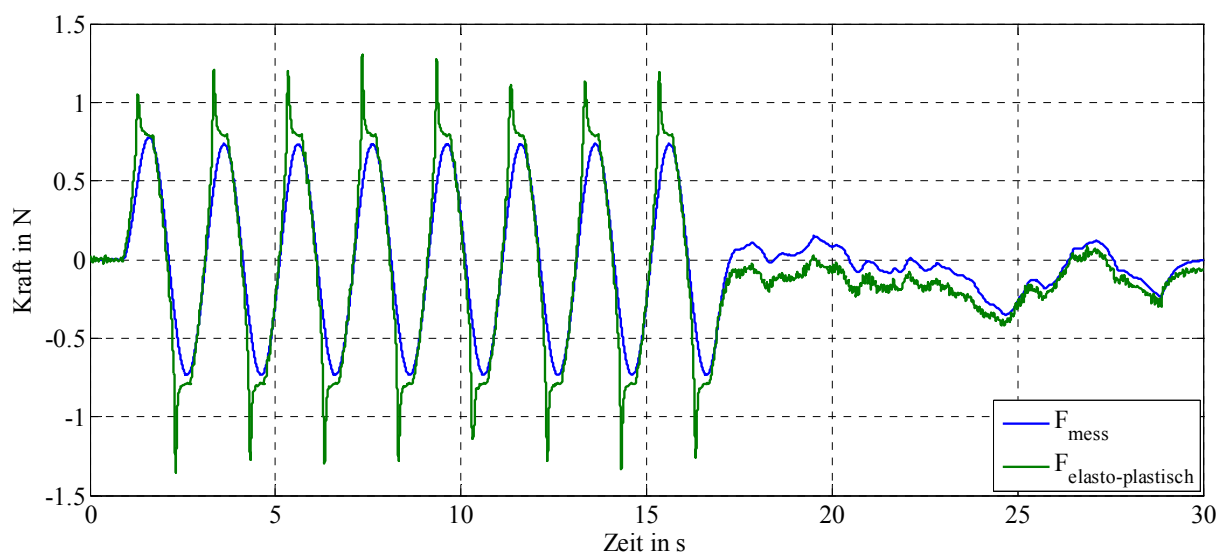


Abbildung 3.25: Vergleich zwischen erregender Kraft und geschätzter Kraft

Somit sind die Ergebnisse des elasto-plastischen Modells leicht besser als die des LuGre-Modells, aber es sind bei beiden Modellen deutliche strukturelle Schwächen zu erkennen. Darüber hinaus kann wieder der bereits erwähnte datensatzabhängige Offset beobachtet werden.

GMS-Reibmodell

Aufgrund der Tatsache, dass das GMS-Modell wie die vorangegangenen Reibmodelle nur die Reibkraft abbildet, wird auch hier zu Beginn der Validierungsdatensatz um die Trägheitskraft bereinigt. Die nachfolgend ermittelten Validierungsergebnisse sind in Abbildung 3.26 dargestellt. Dieser Modellansatz ist ebenfalls in der Lage, das globale Reibverhalten sehr gut nachzubilden. Aber auch hier ist wieder ein datensatzabhängiger Offset klar zu erkennen und so erreicht das Modell lediglich einen relativ großen RMSE von 0,136 N.

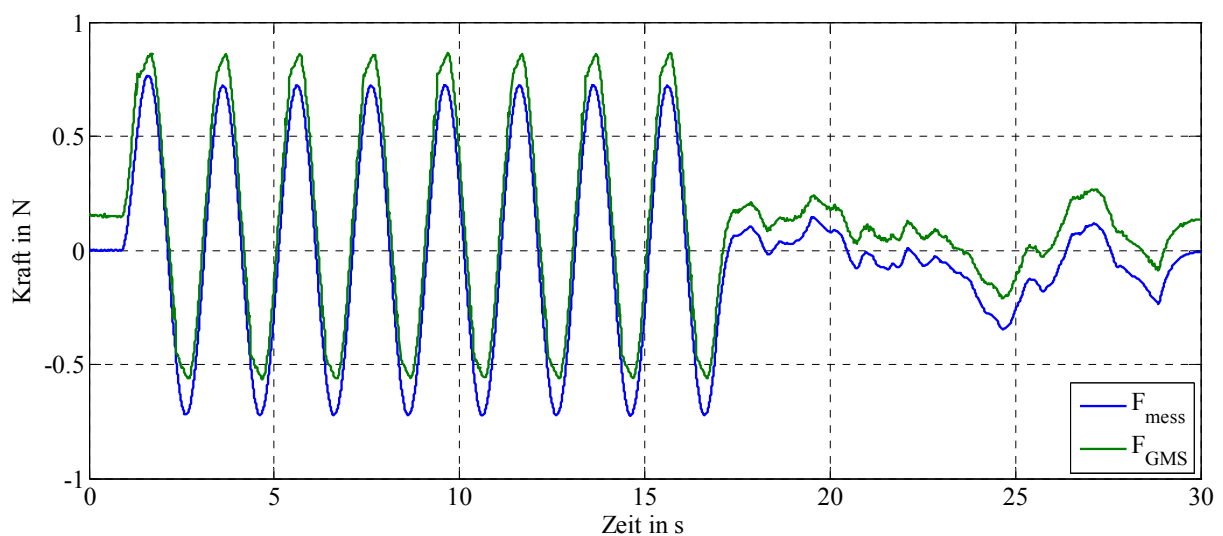


Abbildung 3.26: Vergleich zwischen erregender Kraft und geschätzter Kraft

DNLRX-Reibmodell

In Abbildung 3.27 zeigt die Validierungsergebnisse des identifizierten DNLRX-Modells.

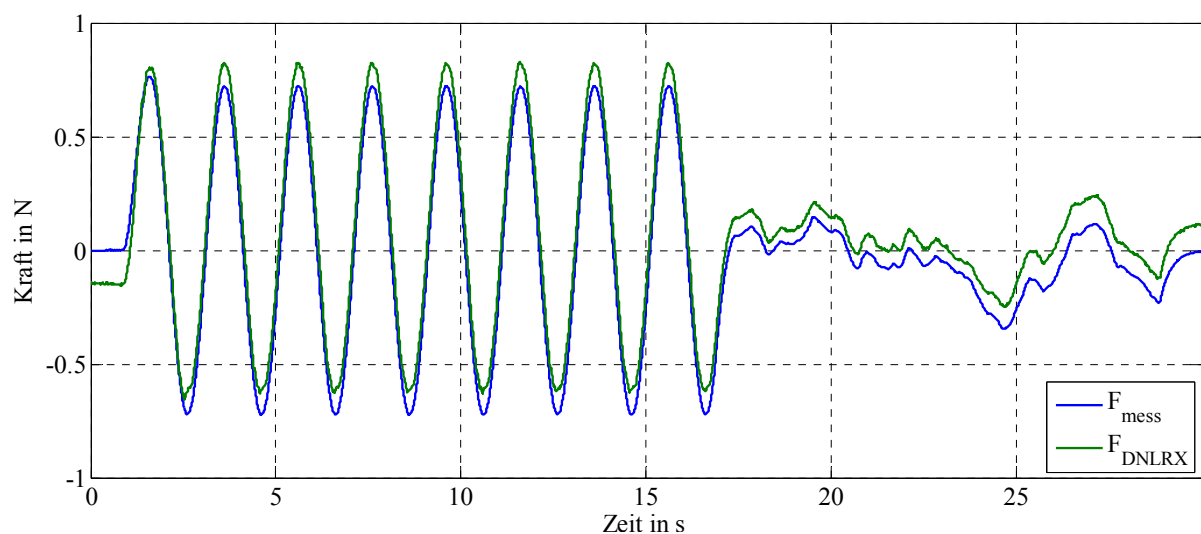


Abbildung 3.27: Vergleich zwischen erregender Kraft und geschätzter Kraft

Das DNLRX-Modell bildet neben der Reibkraft auch die Trägheit des Systems ab und so konnte der Datensatz aus Abbildung 3.23 ohne Modifikationen Anwendung finden. Der gemessene Kraftverlauf wird durch das DNLRX-Modell bis auf einen datensatzabhängigen Offset sehr gut nachgebildet und damit wird ein RMSE von 0,0934 N erzielt.

Künstliches neuronales Netz

Im Anschluss an das Training wird anhand des Validierungsdatensatzes aus Abbildung 3.23 die Generalisierungsfähigkeit des künstlichen neuronalen Netzes überprüft. Die Ergebnisse hierzu sind in Abbildung 3.28 dargestellt. Das KNN bildet das globale Reibverhalten zufriedenstellend nach und der RMSE beträgt 0,091 N. In Analogie zu den vorangegangenen Modellen resultiert auch hier der Großteil des Fehlers aus einem datensatzabhängigen Offset im geschätzten Kraftverlauf.

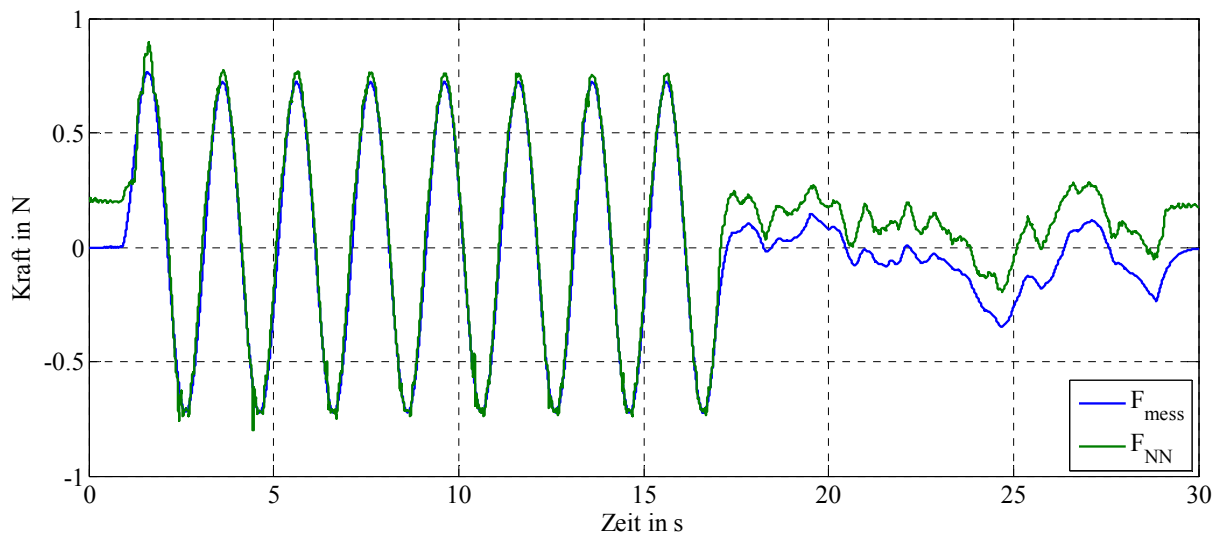


Abbildung 3.28: Vergleich zwischen erregender Kraft und geschätzter Kraft

3.4.4.2 Leistungsbewertung

Zusammenfassend kann festgestellt werden, dass alle untersuchten Reibmodelle in der Lage sind, das Reibverhalten auf Nanometerebene nachzubilden. Am besten schnitten bei der Validierung das DNLRX-Reibmodell und das künstliche Neuronale Netz ab.

Basierend auf diesen Ergebnissen werden im weiteren Verlauf der vorliegenden Arbeit das DNLRX-Modell sowie das künstliche neuronale Netz zur Kompensation der Nichtlinearitäten im System herangezogen. Grund für diese Entscheidung ist der Umstand, dass beide Modellansätze die besten Validierungsergebnisse gezeigt haben. Darüber hinaus bilden beide die gesamte Systemdynamik ab und somit entfällt eine zusätzliche Modellierung der Trägheitskräfte.

Weiterhin zeigte sich, dass die identifizierten Modelle zwar den Verlauf des Validierungsdatensatzes zufriedenstellend abbilden können, aber immer einen nicht unerheblichen Offsetfehler erzeugen. Diese Beobachtung lässt auf ein stark ortsvariantes Systemverhalten schließen, denn die Vorhersagegüte der Modelle korreliert stark mit der Position der betrachteten Bewe-

gungsachse. Die Ursachen für dieses Verhalten sind sowohl stochastischer als auch systematischer Natur. Zu nennen wären in diesem Zusammenhang Schmutzpartikel und Produktionstoleranzen in den Kugelführungen, die Elastizität der Flexbandkabel, die Position der Kugelkäfige, positionsabhängige Lastmomente in den Kugellagern, et cetera. Des Weiteren konnte durch Langzeitexperimente gezeigt werden, dass sich das Systemverhalten nicht nur ortsvariant, sondern auch zeitvariant verhält. Dies resultiert primär aus thermisch bedingten Verformungen im mechanischen Aufbau. Diese Beobachtungen führen zu der Erkenntnis, dass es zu einer effektiven Vorsteuerung des inversen Systemverhaltens notwendig sein wird, die Modellparameter in Echtzeit nachzuführen. Nur so kann der Orts- und Zeitvarianz des Systemverhaltens aktiv entgegengewirkt werden. Beide ausgewählte Modelle bieten hierzu beste Voraussetzungen.

4 Dynamisches Regelungskonzept

Wie bereits in der Zielstellung der vorliegenden Arbeit definiert wurde (siehe Abschnitt 0), soll die Positionierdynamik des in Kapitel 2 vorgestellten Nanopositionier- und Nanomesssystems deutlich gesteigert werden. Hierzu müssen moderne Regelstrategien Anwendung finden, und ein aus der Robotik stammender *modellbasierter Regelungsansatz* bietet hierbei einen möglichen Zugang [31, 34, 35, 91]. Die Struktur dieses Konzeptes ist schematisch in Abbildung 4.1 dargestellt.

Prinzipiell besteht die abgebildete Trajektorienfolgeregelung aus einem Bahnplanungsalgorithmus, einer Vorsteuerung, einem Regler sowie einem Störbeobachter. Die Generierung der dynamischen Sollwerte erfolgt durch den Bahnplanungsalgorithmus, der neben der eigentlichen Bahn- und Trajektorienplanung auch die Synchronisation der Bewegungsachsen gewährleistet. Im Rahmen dieser Vorabplanung werden für jede Achse des NPM-Systems die kinematischen Trajektorien für Weg, Geschwindigkeit, Beschleunigung, Ruck und Ruckanstieg derart geplant, dass eine beliebige Bewegung des Tool Center Points (TCP) im dreidimensionalen Raum umgesetzt werden kann. Die Vorsteuerung generiert in Kenntnis des Streckenmodells genau die Stellgröße, mit der das System im Idealfall auch ungeregelt der Soll-Trajektorie folgen würde. Treten dennoch Fehler zwischen Soll- und Ist-Trajektorie auf, greift eine unterlagerte Regelung ein und kompensiert die Modellungenauigkeiten zwischen Vorsteuerung und realem System sowie gegebenenfalls exogene Störeinflüsse. Zur weiteren Steuerung der Systemdynamik kommt zusätzlich zum inversen Systemmodell in der Vorsteuerung ein Störgrößenbeobachter zum Einsatz. Dieser schätzt nicht messbare Störungen, welche beispielsweise durch stochastische Phänomene in den Kugelführungen (z.B. Verschmutzungen) oder sich ändernde Umweltbedingungen (Temperatur-, Luftfeuchteschwankungen, etc.) verursacht werden. Im Gegensatz zum Bahnplanungsalgorithmus, welcher Sollwerte für alle Bewegungsachsen berechnet, verfügt jede Bewegungsachse über die unabhängigen Regelungskomponenten Vorsteuerung, unterlagerte Regelung und Störgrößenbeobachter.

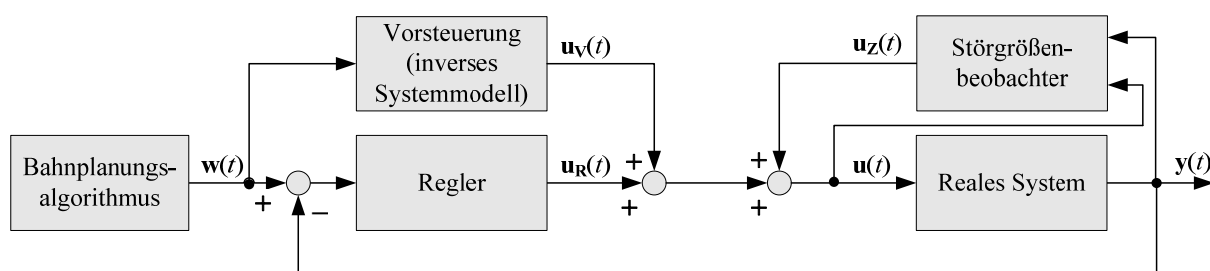


Abbildung 4.1: Konzeption einer Trajektorienfolgeregelung

In diesem Kapitel werden die erwähnten Bestandteile einer Trajektorienfolgeregelung für NPM-Systeme vorgestellt und deren Funktionsweise ausführlich erläutert.

4.1 Bahnplanung

Dieser Abschnitt stellt den in [7, 82, 130, 131] entwickelten Bahnplanungsalgorithmus vor. Nach einer Begriffsbestimmung werden die mathematischen Grundlagen zur Beschreibung von Raumkurven eingeführt. Anschließend wird die Darstellbarkeit der Basiselemente Strecke und Kreisbogen in Abhängigkeit des Bahnparameters s gezeigt und die Restriktionsprojektion auf Bahnebene skizziert. Darauf aufbauend erfolgt die Herleitung einer Bahnplanungsmethode vierter Ordnung, welche stetig differenzierbare Weg-, Geschwindigkeits-, Beschleunigungs- und Ruckprofile generiert. Am Ende dieses Abschnittes wird dargelegt, wie eine kontinuierliche Bahnplanung über mehrere Basiselemente hinweg realisiert werden kann.

4.1.1 Begriffsbestimmung

Im Rahmen der Begriffsbestimmung ist es zunächst unerlässlich, die Bezeichnungen **Bahn** und **Trajektorie** näher zu erläutern. Im Verlauf der vorliegenden Arbeit wird häufig zwischen eindimensionalen und mehrdimensionalen Planungsschritten unterschieden. Um eine einheitliche und nachvollziehbare Kennzeichnung zu erreichen, wird in den folgenden Abschnitten der Begriff **Bahn** gleichbedeutend mit eindimensionalen Zusammenhängen gebraucht. Funktionsverläufe im kartesischen Koordinatensystem werden hingegen als **Kurve** oder **Trajektorie** bezeichnet und die daraus resultierenden dreidimensionalen Größen werden in Vektorschreibweise angegeben. Trotz dieser klaren Unterscheidung liegt eine enge inhaltliche Verwandtschaft dieser Begriffe vor. Der Unterschied liegt ausschließlich in der Wahl des jeweiligen Bezugssystems. Während die **Trajektorie** die eigentliche **Kurve** im kartesischen Koordinatensystem darstellt, entspricht die **Bahn** der eindimensionalen Strecke entlang dieser Raumkurve¹³.

Neben dieser grundlegenden Definition spielen in den folgenden Abschnitten die Begriffe Ort, Geschwindigkeit, Beschleunigung, Ruck und Ruckanstieg als kinematische Größen eine zentrale Rolle. Mathematisch betrachtet, stellt der Ort (eigentlich die Ortsfunktion) eine Funktion $s(t)$ bzw. $\vec{s}(t)$ dar, die jeder Zeit t einen Ort s als Funktionswert zuordnet. In günstigen Fällen kann diese Funktion mehrfach nach der Zeit abgeleitet werden¹⁴. Die einzelnen Ableitungen werden wie allgemein üblich als Geschwindigkeit $v(t)$ bzw. $\vec{v}(t)$, Beschleunigung $a(t)$ bzw. $\vec{a}(t)$ und Ruck $j(t)$ bzw. $\vec{j}(t)$ bezeichnet. Da die in [130] zusätzlich eingeführte vierte

¹³ Auf der so genannten **Bahnebene** gehen sämtliche Richtungsinformationen verloren.

¹⁴ Da das später skizzierte Verfahren auf einer mehrfachen Integration der höheren Ableitungen von $s(t)$ bzw. $\vec{s}(t)$ basiert, kann die mehrfache Differenzierbarkeit als gegeben angesehen werden.

Ableitung von $s(t)$ bzw. $\vec{s}(t)$ in der Fachliteratur nicht einheitlich benannt ist, wurde hierfür der Begriff „Ruckanstieg“ und das mathematische Formelzeichen $d(t)$ bzw. $\vec{d}(t)$ gewählt.

In mechatronischen Systemen können die genannten kinematischen Größen konstruktionsbedingt nur bestimmte Maximalwerte annehmen. Die Limitierung kann unter anderem damit begründet werden, dass die physikalischen Eigenschaften der mechanischen und elektrischen Komponenten eben diese Begrenzungen direkt oder indirekt implizieren. So entspricht die Beschleunigung einer Kraft, die auf eine Masse wirkt. Wird diese Kraft durch einen elektrischen Antrieb aufgebracht, korreliert sie mit dem Strom der durch den Motor fließt. Da aus technischer Sicht die Stromstärke nur bestimmte maximale Werte annehmen und auch nicht sprungförmig verändert werden kann, resultiert daraus eine Limitierung von Beschleunigung und Ruck.

Wichtiger erscheint im Kontext der Erarbeitung neuer Konzepte für NPM-Systeme aber die Tatsache, dass der qualitative und quantitative zeitliche Verlauf der Kinematikfunktionen direkten Einfluss auf die Genauigkeit von Positionierung und Messung nimmt. So verweisen zum Beispiel [41] und [94] darauf, dass durch unkontrollierte Ruckprofile Resonanzen im mechanischen Aufbau des Systems angeregt werden können. Im günstigsten Fall sind lediglich Beeinträchtigungen der Positioniergenauigkeit zu erwarten, im Bereich der Anregung mit Frequenzen nahe der Eigenfrequenz können jedoch auch Beschädigungen der eingesetzten Bauteile hervorgerufen werden. Die Restriktionen werden im Folgenden mit v_{max} , a_{max} , j_{max} und d_{max} bezeichnet. Diese Maximalwerte begrenzen die zulässigen absoluten Funktionswerte der jeweiligen kinematischen Größen und stellen neben der eigentlichen Trajektorie die zentralen Parameter des Planungsmoduls dar. Abschließend sei darauf hingewiesen, dass in den folgenden Ausführungen auf die Angabe physikalischer Einheiten bzw. deren rechnerische Überprüfung weitestgehend verzichtet wird. Alle auftretenden Größen wurden auf SI-Einheiten normiert.

4.1.2 Grundlagen der analytischen Bahnplanung

Im vorangegangenen Abschnitt wurde zwischen ein- und mehrdimensionalen Planungsschritten unterschieden. Wie in Abbildung 4.2 dargestellt ist, kann das Vorgehen bei der analytischen Bahnplanung in drei Phasen unterteilt werden. Zu Beginn findet eine so genannte Dimensionsreduktion statt. Dies bedeutet, dass die dreidimensionale Trajektorie als Funktion des Bahnortes¹⁵ s dargestellt wird und die kinematischen Restriktionen der einzelnen Bewegungsachsen auf die Bahn projiziert werden. Anschließend erfolgt die Kinematikberechnung des Bahnparameters in einer Dimension. Unter Ausnutzung des analytischen Zusammenhangs zwischen dem Bahnparameter und der dreidimensionalen Trajektorie werden im dritten Schritt die kinematischen Trajektorien der Bewegungsachsen bestimmt. Dieses Vorgehen bietet den entscheidenden Vorteil, dass sich die Teilbewegungen der Maschinenachsen zwangsläufig synchron zueinander verhalten.

¹⁵ Der Bahnort ist gleichzeitig der Kurvenparameter. Beide Größen werden im Folgenden mit s bezeichnet.

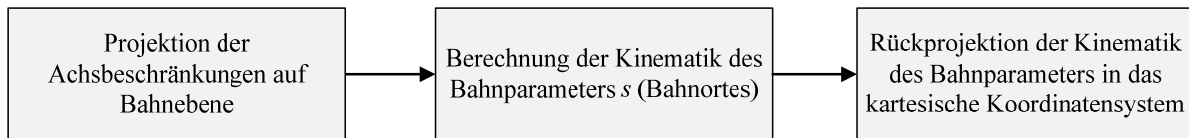


Abbildung 4.2: Prinzipielle Vorgehensweise bei der analytischen Bahnplanung

In diesem Abschnitt sollen nun die mathematischen Grundlagen gelegt werden, um die Dimensionsreduktion sowie die Rückprojektion von einer eindimensionalen Bahn auf die Bewegungstrajektorien der Maschinenachsen durchführen zu können. Der dazwischen liegende Schritt der Kinematikplanung auf Bahnebene erfolgt in Abschnitt 4.1.4.

4.1.2.1 Zusammenhang zwischen Bahn- und Trajektorienkinematik

Um das beschriebene Konzept der analytischen Bahnplanung umzusetzen, muss die Möglichkeit bestehen, von der Bahnkinematik auf die Kinematik der Trajektorie zu schließen. Im Allgemeinen kann diese Umrechnung von Bahn- zur Kurvenkinematik nur dann erfolgen, wenn ein Zusammenhang zwischen dem Ort im kartesischen Koordinatensystem und einem so genannten Kurvenparameter s gefunden werden kann:

$$\vec{s} = \vec{s}(s) \quad (4.1)$$

Wie in [89] geschildert wird, bildet diese Beziehung die zentrale Grundlage zur analytischen Untersuchung räumlicher Kurven, denn nun kann die folgende allgemeine Betrachtung Anwendung finden.

Ist die Trajektorie in \mathbb{R}^3 nach Gleichung (4.1) darstellbar, so kann aus der vektorwertigen Ortsfunktion das so genannte Frenetsche Dreiein inklusive der zugehörigen skalaren Größen Krümmung und Torsion hergeleitet werden. Diese Orthonormalbasis (siehe Abbildung 4.3) bildet das allgemeine Fundament zur geschlossenen Beschreibung räumlicher Kurven und wird daher im Folgenden zunächst eingeführt. Den Ausführungen in [18] folgend, kann zu jedem Punkt einer Trajektorie im \mathbb{R}^3 eine orthonormale Basis aus Tangenten-, Normalen- und Binormalenvektor angegeben werden. Die Grafik zeigt diese drei Vektoren, aus denen sich das frenetsche Dreiein zusammensetzt, für einen frei gewählten Punkt auf der abgebildeten Raumkurve. Der Tangentenvektor \vec{t} liegt tangential an der Trajektorie und gibt die momentane Bewegungsrichtung für den untersuchten Kurvenpunkt an. Der Normalenvektor \vec{n} steht senkrecht auf \vec{t} und zeigt in die Richtung des Drehzentrums des so genannten Schmiegekreises¹⁶. Geometrisch betrachtet, gibt der Normalenvektor somit die Richtung an, in die der Tangentenvektor gedreht werden muss, wenn die Bewegung entlang der Trajektorie fortgesetzt werden soll. Der Binormalenvektor \vec{b} vervollständigt die Orthonormalbasis und verkörpert den Normalenvektor der Ebene, in der sich der Schmiegekreis befindet.

¹⁶ Vereinfacht entspricht der Schmiegekreis exakt dem Kreis, der die Kurve im untersuchten Punkt berührt bzw. schneidet und die Trajektorie in der Umgebung des Punktes am besten annähert.

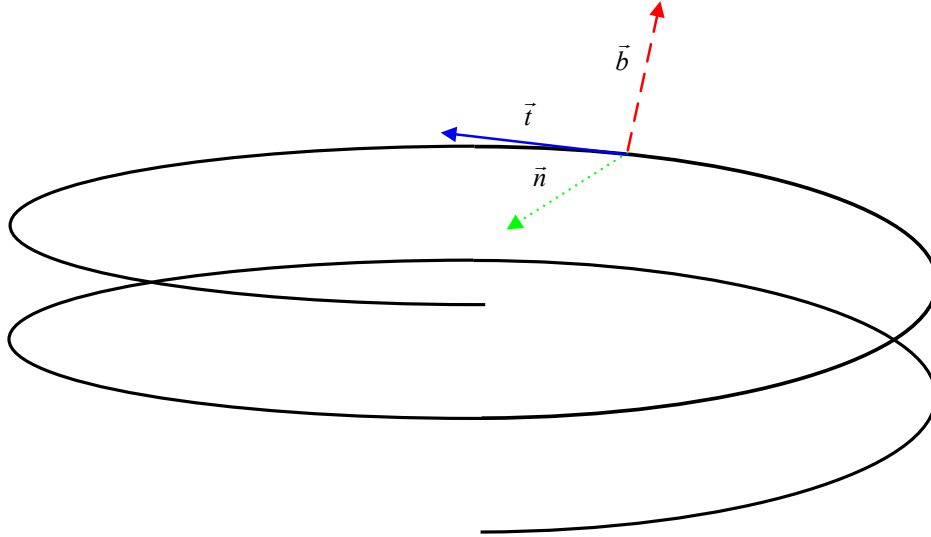


Abbildung 4.3: Trajektorie mit frenetschem Dreiein

Da der Tangentenvektor die Änderung der Position in Abhängigkeit der zurückgelegten Strecke angibt, kann er als Ableitung der Kurvenfunktion nach s dargestellt werden:

$$\vec{t} = \vec{t}(s) = \vec{s}' = \frac{d\vec{s}}{ds} \quad (4.2)$$

Der Normalenvektor beschreibt die Änderung der Tangentenrichtung pro zurückgelegten Bahnweg und wird daher durch Ableiten des Tangentenvektors nach dem Bahnort s und anschließender Normierung gewonnen:

$$\vec{n} = \vec{n}(s) = \frac{\vec{t}'}{|\vec{t}'|} \quad (4.3)$$

Aus dem Tangenten- und dem Normalenvektor kann der Binormalenvektor durch Berechnung des Kreuzproduktes bestimmt werden:

$$\vec{b} = \vec{b}(s) = \vec{t} \times \vec{n} \quad (4.4)$$

Während der Tangentenvektor die Richtung der Bewegung angibt, stellen der Normalen- und der Binormalenvektor die beiden Freiheitsgrade dar, in welchen sich die Tangentenrichtung ändern kann. Da dieser qualitative Zusammenhang nicht ausreicht, um eine tatsächliche Änderung auch quantitativ beschreiben zu können, müssen die Freiheitsgrade um die Koeffizienten Krümmung κ und Torsion τ ergänzt werden. Da die Krümmung den Betrag der Richtungsänderung der Tangente in Abhängigkeit der Bahnstrecke angibt, gilt:

$$\kappa = \kappa(s) = |\vec{t}'| \quad (4.5)$$

In Analogie gibt die Torsion den Betrag der Verkipfung der Schmiegekreisebene an und kann mittels folgender Vorschrift gewonnen werden:

$$\tau = \tau(s) = |\vec{b}'| \quad (4.6)$$

Neben dieser grundlegenden Definition der Parameter können die Ableitungen¹⁷ der einzelnen Basisvektoren auch durch Linearkombinationen der bisher eingeführten Vektoren dargestellt werden. Diese so genannten Frenetschen Formeln können wie folgt angegeben werden:

$$\begin{aligned}\vec{t}' &= \kappa \cdot \vec{n} \\ \vec{n}' &= \tau \cdot \vec{b} - \kappa \cdot \vec{t} \\ \vec{b}' &= -\tau \cdot \vec{n}\end{aligned}\tag{4.7}$$

Die Begründung, dass das frenetsche Dreibein die Grundlage für eine geschlossene Beschreibung räumlicher Kurven darstellt, liefert die Herleitung der Trajektorienkinematik. In Analogie zum eindimensionalen Fall geht auch für eine vektorwertige Ortsfunktion die Geschwindigkeit aus deren zeitlicher Ableitung hervor. Die Differentiation von Gleichung (4.1) nach der Zeit ergibt den folgenden Zusammenhang:

$$\vec{v} = \dot{\vec{s}} = \frac{d\vec{s}}{ds} \cdot \frac{ds}{dt} = \vec{s}' \cdot v = \vec{t} \cdot v\tag{4.8}$$

Wird Gleichung (4.8) erneut nach der Zeit abgeleitet, kann die Berechnungsvorschrift für die Kurvenbeschleunigung gewonnen werden:

$$\vec{a} = \dot{\vec{v}} = \dot{\vec{t}} \cdot v + \vec{t} \cdot \dot{v} = \vec{t}' \cdot \dot{s} \cdot v + \vec{t} \cdot \dot{v} = \vec{n} \cdot \kappa \cdot v^2 + \vec{t} \cdot a\tag{4.9}$$

Entsprechend führen zwei weitere Differentiationsschritte zu den Berechnungsvorschriften der vektorwertigen Funktionen des Kurvenrucks j und des Kurvenruckanstiegs d :

$$\vec{j} = \vec{b} \kappa \tau v^3 + \vec{n} \cdot (\kappa' v^3 + 3\kappa v a) + \vec{t} \cdot (j - \kappa^2 v^3)\tag{4.10}$$

$$\begin{aligned}\vec{d} &= \vec{b} \cdot (6\kappa \tau v^2 a + 2 \cdot \kappa' \tau v^4 + \kappa \tau' v^4) + \\ &\quad \vec{n} \cdot (3\kappa a^2 + 4\kappa v j - \kappa^3 v^4 - \kappa \tau^2 v^4 + 6\kappa' v^2 \cdot a + \kappa'' v^4) + \\ &\quad \vec{t} \cdot (d - 6\kappa^2 v^2 a - 3\kappa \kappa' v^4)\end{aligned}\tag{4.11}$$

Durch die Gleichungen (4.8) bis (4.11) stehen nun vektorwertige Funktionen für alle untersuchten kinematischen Größen zur Verfügung. Mittels dieser Berechnungsvorschriften kann zu jedem Zeitpunkt durch Einsetzen der durch die Bahnplanung (siehe Abschnitt 4.1.4) ermittelten skalaren Bahngrößen v, a, j, d die Umrechnung in die Kinematik der Trajektorie erfolgen. Im folgenden Abschnitt werden die hergeleiteten Beziehungen nun auf die dreidimensionalen Raumkurven Strecke und Kreisbogen angewendet, da der vorgestellte Bahnplanungsalgorithmus diese Trajektorien als Basiselemente nutzt.

4.1.2.2 Bestimmung der Kurvenkinematik für die Strecke

Eine beliebige Strecke kann durch einen Start- und einen Endpunkt festgelegt werden. Wie aus der Geometrie bekannt ist, kann die resultierende Gerade im \mathbb{R}^3 folgendermaßen dargestellt werden:

¹⁷ nach dem Bahnort s

$$\vec{s} = \vec{s}_{start} + \lambda \cdot (\vec{s}_{ziel} - \vec{s}_{start}) \quad (4.12)$$

Nach Gleichung (4.12) ist der Ortsvektor jedes auf dieser Geraden liegenden Punktes \vec{s} darstellbar als Summe eines so genannten Aufpunktes \vec{s}_{start} und einem Vielfachen λ des Richtungsvektors $(\vec{s}_{ziel} - \vec{s}_{start})$. Der notwendige Zusammenhang zwischen Bahn- und Trajektorienplanung kann also hergestellt werden, wenn der Bahnort s in Abhängigkeit des Kurvenparameters λ ausgedrückt werden kann. Es ist leicht nachvollziehbar, dass der zu $s = s_{start} = 0$ korrespondierende Wert für λ ebenfalls 0 ist. Der Endpunkt der Strecke entspricht einem λ von 1 und einem Bahnweg von \vec{s}_{ziel} . Eine Division durch \vec{s}_{ziel} liefert die notwendige Neuskalierung des Wertebereichs für s . Somit kann Gleichung (4.12) in die folgende Darstellung überführt werden:

$$\vec{s} = \vec{s}_{start} + \frac{s}{|\vec{s}_{ziel} - \vec{s}_{start}|} \cdot (\vec{s}_{ziel} - \vec{s}_{start}) \quad (4.13)$$

Mit Gleichung (4.13) ist nun eine Beziehung zwischen dem Bahnort s und der Strecke \vec{s} hergeleitet worden. Somit kann zu jedem Zeitpunkt ein durch die Bahnplanung vorgegebener Bahnort in den korrespondierenden Punkt im \mathbb{R}^3 umgerechnet werden.

Mit Gleichung (4.13) ist die Voraussetzung geschaffen worden, die Frenetschen Formeln für das Basiselement Strecke anzuwenden. Durch einmalige Ableitung der Gleichung (4.13) nach dem Bahnort ergibt sich:

$$\vec{t} = \vec{s}' = \frac{\vec{s}_{ziel} - \vec{s}_{start}}{|\vec{s}_{ziel} - \vec{s}_{start}|} = \vec{r}_s \quad (4.14)$$

Es ist zu sehen, dass der resultierende Tangentialvektor \vec{t} eine konstante vektorwertige Funktion darstellt. Hier wird deutlich, dass die Tangente mit dem normierten Richtungsvektor der Strecke übereinstimmt, der im Folgenden verkürzt als \vec{r}_s bezeichnet werden soll. Aus der Funktion des Tangentenvektors kann durch erneute Differentiation der Normalenvektor bestimmt werden. Da die Tangente als konstante Funktion keinerlei Abhängigkeit von s aufweist, führt dieser Ableitungsschritt zu einem Nullvektor. Auch die anschließende Berechnung des Kreuzproduktes aus \vec{t} und \vec{n} hat einen Nullvektor zum Ergebnis. Dies kann wie folgt zusammengefasst werden:

$$\vec{n} = \vec{t}' = \vec{r}_s' = \vec{0} \quad (4.15)$$

$$\vec{b} = \vec{t} \times \vec{n} = \vec{r}_s \times \vec{0} = \vec{0} \quad (4.16)$$

Unter Kenntnis dieser drei Vektoren kann auch die Berechnung der Krümmung und der Torsion durchgeführt werden:

$$\kappa = |\vec{t}'| = |\vec{r}_s'| = |\vec{0}| = 0 \quad (4.17)$$

$$\tau = |\vec{b}'| = |\vec{0}'| = |\vec{0}| = 0 \quad (4.18)$$

Durch Einsetzen der Gleichungen (4.14) bis (4.18) in die Berechnungsvorschriften (4.8) bis (4.11) können die folgenden Zusammenhänge ermittelt werden:

$$\vec{v} = \vec{r}_s \cdot v \quad (4.19)$$

$$\vec{a} = \vec{r}_s \cdot a \quad (4.20)$$

$$\vec{j} = \vec{r}_s \cdot j \quad (4.21)$$

$$\vec{d} = \vec{r}_s \cdot d \quad (4.22)$$

Hier ist zu sehen, dass für geradlinige Kurvenformen aufgrund des Wegfalls von Normalenvektor, Binormalenvektor, Krümmung und Torsion die Trajektorienkinematik durch eine einfache Multiplikation der jeweiligen Bahngröße mit dem normierten Richtungsvektor berechnet werden kann.

4.1.2.3 Bestimmung der Kurvenkinematik für den Kreisbogen

In ähnlicher Art und Weise kann auch für den Kreisbogen als weitere einfache Basiskurve eine Umrechnung erfolgen. Grundsätzlich legt folgende Gleichung einen Kreis im \mathbb{R}^3 fest:

$$\vec{s} = \vec{s}_{mitte} + \mathbf{R}(\varphi) \cdot (\vec{s}_{start} - \vec{s}_{mitte}) \quad (4.23)$$

Es ist zu erkennen, dass jeder Punkt \vec{s} auf dem Kreisbogen durch eine Rotation \mathbf{R} des Radiusvektors $(\vec{s}_{start} - \vec{s}_{mitte})$ um einen Winkel φ und anschließende Addition des Kreismittelpunktes \vec{s}_{mitte} bestimmt werden kann. Da die Rotationsachse eines Kreisbogens konstant bleibt, ist \mathbf{R} nur vom Drehwinkel φ abhängig, der somit als Kurvenparameter angesehen werden kann. Die Umrechnung zwischen dem Bahnort s auf dem Kreisbogen und dem Kurvenparameter φ kann durch einen einfachen Dreisatz erfolgen und führt zu folgendem Zusammenhang:

$$\vec{s} = \vec{s}_{mitte} + \mathbf{R}\left(\frac{s}{|\vec{s}_{start} - \vec{s}_{mitte}|} \cdot \varphi_{ziel}\right) \cdot (\vec{s}_{start} - \vec{s}_{mitte}) \quad (4.24)$$

Die Rotationsmatrix des \mathbb{R}^3 wird durch eine normierte Drehachse $\vec{r}_a = (r_{a,x} \ r_{a,y} \ r_{a,z})^T$ und einen Drehwinkel φ wie folgt berechnet:

$$\mathbf{R}(\varphi) = \begin{pmatrix} a_\varphi + r_{a,x}^2 a_\varphi & r_{a,x} r_{a,y} a_\varphi - r_{a,z} \sin \varphi & r_{a,x} r_{a,z} a_\varphi + r_{a,y} \sin \varphi \\ r_{a,x} r_{a,y} a_\varphi + r_{a,z} \sin \varphi & \cos \varphi + r_{a,y}^2 a_\varphi & r_{a,y} r_{a,z} a_\varphi - r_{a,x} \sin \varphi \\ r_{a,x} r_{a,z} a_\varphi - r_{a,y} \sin \varphi & r_{a,y} r_{a,z} a_\varphi + r_{a,x} \sin \varphi & \cos \varphi + r_{a,z}^2 a_\varphi \end{pmatrix} \quad (4.25)$$

$$\text{mit } a_\varphi \triangleq (1 - \cos \varphi)$$

Nun können anhand der Gleichung (4.24) nach Einsetzen von (4.25) die einzelnen Differentiationsschritte zur Berechnung der Frenetschen Formeln durchgeführt werden. Da es sich beim Kreisbogen aber um eine ebene Raumkurve handelt, kann durch eine Koordinatentransformation der Berechnungsaufwand erheblich reduziert werden. Wie beispielsweise in [82] erläutert wird, kann die Berechnung einer ebenen Trajektorie in \mathbb{R}^3 durch entsprechende Transformationen auf einen Entwurf im \mathbb{R}^2 reduziert werden. Das Ergebnis kann anschließend wiederum mittels der inversen Transformation in die korrekte Lage im \mathbb{R}^3 projiziert werden.

Die Basistransformation kann wie folgt formuliert werden:

$$\tilde{\vec{s}} = \mathbf{T}_r \cdot (\vec{s} + \vec{t}_t) \quad (4.26)$$

Somit ergeben sich die folgenden transformierten Kreisparameter:

$$\tilde{\vec{s}}_{mitte} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (4.27)$$

$$\tilde{\vec{s}}_{start} = \begin{pmatrix} |\vec{s}_{start} - \vec{s}_{mitte}| \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix} \quad (4.28)$$

$$\tilde{\vec{r}}_a = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (4.29)$$

Der Umstand, dass es sich bei den neuen Größen um Transformierte der ursprünglich gegebenen Vektoren handelt, soll die Tilde über den Bezeichnern verdeutlichen. Nach der erfolgten Transformation liegt nun der Kreismittelpunkt im Ursprung, während die Drehachse mit der \tilde{z} -Achse zusammenfällt. Der Startpunkt der Drehung liegt auf der \tilde{x} -Achse und die Länge des Radiusvektors $|\vec{s}_{start} - \vec{s}_{mitte}|$ wird nunmehr mit dem Formelzeichen r benannt.

Durch die Transformation ergibt sich für die Funktion des Kurvenorts:

$$\tilde{\vec{s}} = \tilde{\mathbf{R}} \left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel} \right) \cdot \tilde{\vec{s}}_{start} \quad (4.30)$$

und für die transformierte Rotationsmatrix:

$$\tilde{\mathbf{R}}(\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (4.31)$$

Durch Ableitung der Gleichung (4.30) unter Nutzung von (4.28) kann der Tangentenvektor im transformierten Koordinatensystem ermittelt werden:

$$\tilde{\vec{t}} = \tilde{\vec{s}}' = \frac{r \cdot \varphi_{ziel}}{s_{ziel}} \cdot \begin{pmatrix} -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ \cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix}, \quad \text{mit} \quad \frac{r \cdot \varphi_{ziel}}{s_{ziel}} = 1 \quad (4.32)$$

Ein weiterer Differentiationsschritt liefert den Normalenvektor des Kreisbogens im transformierten Koordinatensystem:

$$\tilde{\vec{n}} = \frac{\tilde{\vec{t}}'}{|\tilde{\vec{t}}'|} = \frac{\frac{\varphi_{\text{ziel}}}{s_{\text{ziel}}} \cdot \begin{pmatrix} -\cos\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ -\sin\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ 0 \end{pmatrix}}{\frac{\varphi_{\text{ziel}}}{s_{\text{ziel}}}} = \begin{pmatrix} -\cos\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ -\sin\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ 0 \end{pmatrix} \quad (4.33)$$

Durch Berechnung des Kreuzproduktes aus $\tilde{\vec{t}}$ und $\tilde{\vec{n}}$ kann auch der Binormalenvektor $\tilde{\vec{b}}$ bestimmt werden:

$$\tilde{\vec{b}} = \tilde{\vec{t}} \times \tilde{\vec{n}} = \begin{pmatrix} -\sin\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ \cos\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ 0 \end{pmatrix} \times \begin{pmatrix} -\cos\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ -\sin\left(\frac{s}{s_{\text{ziel}}} \cdot \varphi_{\text{ziel}}\right) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (4.34)$$

Auf Basis der Gleichungen (4.32) bis (4.34) besteht die Möglichkeit, die folgenden Werte für die Krümmung κ und die Torsion τ zu ermitteln:

$$\kappa = |\tilde{\vec{t}}'| = \frac{\varphi_{\text{ziel}}}{s_{\text{ziel}}} = \frac{\varphi_{\text{ziel}}}{r \cdot \varphi_{\text{ziel}}} = \frac{1}{r} \quad (4.35)$$

$$\tau = |\tilde{\vec{b}}'| = \left| \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}' \right| = |\vec{0}| = 0 \quad (4.36)$$

Die Herleitung liefert ein konstantes κ , das dem Kehrwert des Kreisradius r entspricht und somit exakt mit der allgemeinen Definition der Kreiskrümmung übereinstimmt. Da der Kreisbogen eine ebene Kurve darstellt, findet keine Verkipfung der entsprechenden Ebenennormalen statt und die Torsion entfällt. Es sei an dieser Stelle auch angemerkt, dass die beiden Koeffizienten κ sowie τ skalare Größen darstellen und daher die Transformation des Koordinatensystems auf deren Wert keine Auswirkung hat. Aus diesem Grund behalten beide Größen auch im originalen Koordinatensystem ihre Gültigkeit, was am Fehlen der Tilde zu erkennen ist.

Durch Einsetzen der bestimmten Vektoren des Frenetschen Dreibeins und der Koeffizienten κ und τ in die Gleichungen (4.8) bis (4.11) kann die Trajektorienkinematik für den Kreis angegeben werden:

$$\tilde{\vec{v}} = \begin{pmatrix} -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ \cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix} \cdot v \quad (4.37)$$

$$\tilde{\vec{a}} = \begin{pmatrix} -\cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix} \cdot \frac{v^2}{r} + \begin{pmatrix} -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ \cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix} \cdot a \quad (4.38)$$

$$\tilde{\vec{j}} = \begin{pmatrix} -\cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix} \cdot \frac{3va}{r} + \begin{pmatrix} -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ \cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix} \cdot \left(j - \frac{v^3}{r^2}\right) \quad (4.39)$$

$$\tilde{\vec{d}} = \begin{pmatrix} -\cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix} \cdot \left(\frac{3a^2}{r} + \frac{4vj}{r} - \frac{v^4}{r^3}\right) + \begin{pmatrix} -\sin\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ \cos\left(\frac{s}{s_{ziel}} \cdot \varphi_{ziel}\right) \\ 0 \end{pmatrix} \cdot \left(d - \frac{6v^2a}{r^2}\right) \quad (4.40)$$

Diese vier vektorwertigen Funktionen bilden nun die Berechnungsgrundlage für die kinematischen Größen Geschwindigkeit, Beschleunigung, Ruck und Ruckanstieg des Kreisbogens. Da die Gleichungen (4.37) bis (4.40) jedoch nur im transformierten Koordinatensystem Gültigkeit besitzen, müssen die Ergebnisse mittels der entsprechenden inversen Transformation zurück in das Ausgangskordinatensystem überführt werden:

$$\vec{s} = \mathbf{T}_r^{-1} \cdot \tilde{\vec{s}} - \vec{t}_i \quad (4.41)$$

Im Gegensatz zu den Ortsvektoren bestehen die kinematischen Größen ausschließlich aus Richtungsvektoren, für die nur die rotatorische Modifikation rückgängig gemacht werden muss. Somit kann die translatorische Verschiebung bei der Rücktransformation entfallen und es ergibt sich:

$$\vec{v} = \mathbf{T}_r^{-1} \cdot \tilde{\vec{v}} \quad (4.42)$$

$$\vec{a} = \mathbf{T}_r^{-1} \cdot \tilde{\vec{a}} \quad (4.43)$$

$$\vec{j} = \mathbf{T}_r^{-1} \cdot \tilde{\vec{j}} \quad (4.44)$$

$$\vec{d} = \mathbf{T}_r^{-1} \cdot \tilde{\vec{d}} \quad (4.45)$$

Mithilfe der beschriebenen Berechnungsvorschriften ist eine direkte Umrechnung von Bahnkinematik auf die korrespondierende Trajektorienkinematik für die Strecke und den Kreis möglich.

4.1.3 Bestimmung der Restriktionen auf Bahnebene

Im vorangegangenen Abschnitt wurden Berechnungsvorschriften für die Strecke und den Kreisbogen hergeleitet, die eine analytische Beschreibung der Kurvenkinematik auf Basis der Bahnkinematik ermöglichen. Voraussetzung für die Anwendbarkeit ist eine Darstellung der Trajektorie als eine mindestens dreimalig differenzierbare vektorwertige Funktion des Bahnortes s . Somit kann nach der Kinematikplanung von s auf die entsprechende Kinematik der Trajektorie geschlossen werden. Bevor aber die Kinematikplanung auf Bahnebene begonnen werden kann, müssen die gültigen kinematischen Restriktionen bestimmt werden. Im Folgenden wird diese Projektion der Beschränkungen für die Basiselemente Strecke und Kreisbogen skizziert.

4.1.3.1 Bestimmung der Bahnrestriktionen für die Strecke

Auf Grundlage der Gleichungen (4.19) bis (4.22) kann die Bestimmung der Restriktionen auf Bahnebene erfolgen. Das grundsätzliche Problem hierbei ist die Festlegung von gültigen Intervallen für die kinematischen Bahngrößen, sodass die Achsrestriktionen nicht verletzt werden. Dies kann folgendermaßen veranschaulicht werden:

$$-\vec{v}_{max} \leq \vec{r}_s \cdot \vec{v} \leq \vec{v}_{max} \quad \forall \vec{v} \in [-v_{max}, v_{max}] \quad (4.46)$$

$$-\vec{a}_{max} \leq \vec{r}_s \cdot \vec{a} \leq \vec{a}_{max} \quad \forall \vec{a} \in [-a_{max}, a_{max}] \quad (4.47)$$

$$-\vec{j}_{max} \leq \vec{r}_s \cdot \vec{j} \leq \vec{j}_{max} \quad \forall \vec{j} \in [-j_{max}, j_{max}] \quad (4.48)$$

$$-\vec{d}_{max} \leq \vec{r}_s \cdot \vec{d} \leq \vec{d}_{max} \quad \forall \vec{d} \in [-d_{max}, d_{max}] \quad (4.49)$$

Um den Lösungsweg zu verdeutlichen, sind einige Umformungen nötig, die im Folgenden exemplarisch am Beispiel der Geschwindigkeit gezeigt werden. Zunächst ist es sinnvoll, die Vektorschreibweise in (4.46) durch die explizite Darstellung der Komponenten zu ersetzen und jede Dimension in einer separaten doppelten Ungleichung auszudrücken:

$$-v_{max,x} \leq r_{s,x} \cdot v \leq v_{max,x} \quad \forall v \in [-v_{max}, v_{max}] \quad (4.50)$$

$$-v_{max,y} \leq r_{s,y} \cdot v \leq v_{max,y} \quad \forall v \in [-v_{max}, v_{max}] \quad (4.51)$$

$$-v_{max,z} \leq r_{s,z} \cdot v \leq v_{max,z} \quad \forall v \in [-v_{max}, v_{max}] \quad (4.52)$$

Durch Anwendung der Betragsfunktion ist es möglich, diese drei Ungleichungen zu vereinfachen und nach v umzustellen. Da der Bahnplanungsalgorithmus sicherstellt, dass durch die Parametrierung mit v_{max} die Geschwindigkeitsfunktion das Intervall $[-v_{max}, v_{max}]$ nicht überschreitet, kann v durch die Bahnrestriktion v_{max} ersetzt werden. Durch diese Umformungen

entstehen die folgenden drei oberen Schranken für die maximal zulässige Bahngeschwindigkeit:

$$v_{max} \leq \frac{v_{max,x}}{|r_{s,x}|} \quad (4.53)$$

$$v_{max} \leq \frac{v_{max,y}}{|r_{s,y}|} \quad (4.54)$$

$$v_{max} \leq \frac{v_{max,z}}{|r_{s,z}|} \quad (4.55)$$

v_{max} ist nun der größte Wert, der alle drei Ungleichungen erfüllt:

$$v_{max} = \min \left(\frac{v_{max,x}}{|r_{s,x}|}, \frac{v_{max,y}}{|r_{s,y}|}, \frac{v_{max,z}}{|r_{s,z}|} \right) \quad (4.56)$$

In gleicher Art und Weise können die Bahnrestriktionen für die Beschleunigung, den Ruck und den Ruckanstieg bestimmt werden:

$$a_{max} = \min \left(\frac{a_{max,x}}{|r_{s,x}|}, \frac{a_{max,y}}{|r_{s,y}|}, \frac{a_{max,z}}{|r_{s,z}|} \right) \quad (4.57)$$

$$j_{max} = \min \left(\frac{j_{max,x}}{|r_{s,x}|}, \frac{j_{max,y}}{|r_{s,y}|}, \frac{j_{max,z}}{|r_{s,z}|} \right) \quad (4.58)$$

$$d_{max} = \min \left(\frac{d_{max,x}}{|r_{s,x}|}, \frac{d_{max,y}}{|r_{s,y}|}, \frac{d_{max,z}}{|r_{s,z}|} \right) \quad (4.59)$$

Mittels der Gleichungen (4.56) bis (4.59) ist eine Berechnung der Bahnrestriktionen auf Grundlage des Richtungsvektors der Strecke und der Achsrestriktionen möglich.

4.1.3.2 Bestimmung der Bahnrestriktionen für den Kreisbogen

Analog zum Beispiel der Strecke kann auch für den Kreisbogen die Einhaltung der Achsrestriktionen durch vier doppelte Ungleichungen veranschaulicht werden:

$$-\vec{v}_{max} \leq \mathbf{T}_r^{-1} \cdot \tilde{\vec{v}} \leq \vec{v}_{max} \quad (4.60)$$

$$-\vec{a}_{max} \leq \mathbf{T}_r^{-1} \cdot \tilde{\vec{a}} \leq \vec{a}_{max} \quad (4.61)$$

$$-\vec{j}_{max} \leq \mathbf{T}_r^{-1} \cdot \tilde{\vec{j}} \leq \vec{j}_{max} \quad (4.62)$$

$$-\vec{d}_{max} \leq \mathbf{T}_r^{-1} \cdot \tilde{\vec{d}} \leq \vec{d}_{max} \quad (4.63)$$

Da die inverse Transformationsmatrix \mathbf{T}_r^{-1} eine Rotationsmatrix darstellt, erfordert dies nach Gleichung (4.25) eine Vielzahl trigonometrischer Berechnungen. Durch die Multiplikation mit der transformierten vektorwertigen Kinematik, die ebenfalls Sinus- und Cosinusfunktionen enthält, entstehen für den Kreisbogen wesentlich komplexere Ausdrücke innerhalb der Ungleichungen (4.60) bis (4.63). Darüber hinaus enthalten die einzelnen Ungleichungen nicht

mehr ausschließlich die jeweils zugehörige Bahngröße. Diese beiden Sachverhalte führen dazu, dass es für den Kreisbogen nicht gelingt, mittels einiger Umformungen die Bahnrestriktionen direkt zu bestimmen. Zur Lösung dieses weitaus komplexeren Problems wurde ein alternativer numerischer Lösungsansatz gewählt, der im Folgenden kurz skizziert wird.

Erster Lösungsschritt: Berechnung der Vektorlängen

Die Basis für den angedeuteten Lösungsweg bildet die Tatsache, dass mittels der Gleichungen (4.42) bis (4.45) eine Berechnung der Länge des jeweiligen resultierenden Kinematikvektors (siehe Abbildung 4.4) ohne Anwendung trigonometrischer Funktionen möglich ist¹⁸. Wie bereits erläutert wurde, stehen die Vektoren \vec{r} und \vec{n} senkrecht aufeinander und stellen normierte Basisvektoren dar. Diese werden durch die Koeffizienten in den Gleichungen (4.37) bis (4.40) skaliert. Somit entsteht ein rechtwinkliges Dreieck, in dem die skalierten Basisvektoren die Katheten darstellen, während das Ergebnis der Vektoraddition die Hypotenuse bildet. Unter Nutzung dieses Zusammenhangs ist es möglich, mithilfe des Satzes des Pythagoras die Länge der resultierenden Kinematikvektoren zu bestimmen:

$$|\vec{v}| = \sqrt{v^2} \quad (4.64)$$

$$|\vec{a}| = \sqrt{\left(\frac{v^2}{r}\right)^2 + a^2} \quad (4.65)$$

$$|\vec{j}| = \sqrt{\left(\frac{3va}{r}\right)^2 + \left(j - \frac{v^3}{r^2}\right)^2} \quad (4.66)$$

$$|\vec{d}| = \sqrt{\left(\frac{3a^2}{r} + \frac{4vj}{r} - \frac{v^4}{r^3}\right)^2 + \left(d - \frac{6v^2a}{r^2}\right)^2} \quad (4.67)$$

Zweiter Lösungsschritt: Abschätzung der Vektorlängen

Um aus den Gleichungen (4.64) bis (4.67) eine Aussage über die maximal zulässigen Restriktionen auf Bahnebene zu treffen, muss zunächst eine Abschätzung der oberen Schranke erfolgen. Das heißt, dass für jede kinematische Größe die maximale Vektorlänge bestimmt werden muss, die beliebig in der Kreisebene gedreht werden kann, ohne dabei eine Verletzung der jeweiligen Achsrestriktionen hervorzurufen. Diese Schranken können ermittelt werden, indem zunächst die Schnittgeraden der Kreisebene mit den drei Koordinatenebenen gebildet werden.

Die zugehörigen normierten Richtungsvektoren stellen genau die Richtungen dar, die eine Verletzung der Achsenbeschränkungen hervorrufen. Analog zur Restriktionsbestimmung der Strecke können für jeden dieser Richtungsvektoren obere Schranken für die maximalen Vektorlängen bestimmt werden. Die Auswahl des jeweils kleinsten Wertes liefert daraufhin die maximale Vektorlänge für jede kinematische Größe.

¹⁸ Die Länge des resultierenden Kinematikvektors ist unabhängig vom gewählten Koordinatensystem.

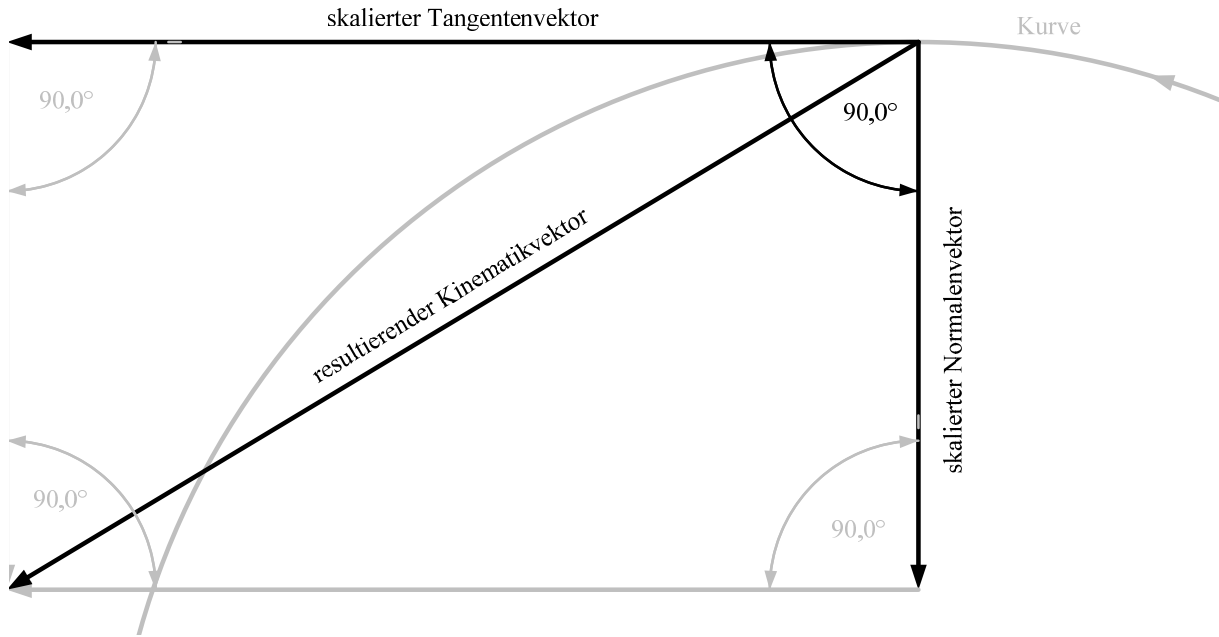


Abbildung 4.4: Rechtwinkliges Dreieck, aus Tangenten-, Normalen- und Kinematikvektor

Daher können die zulässigen Vektorlängen als bekannt angenommen und die Berechnungen in (4.64) bis (4.67) entsprechend umgeformt werden:

$$\sqrt{v^2} \leq \hat{v}_{\max} \quad \forall d \in [-v_{\max}, v_{\max}] \quad (4.68)$$

$$\sqrt{\left(\frac{v^2}{r}\right)^2 + a^2} \leq \hat{a}_{\max} \quad \forall d \in [-a_{\max}, a_{\max}] \quad (4.69)$$

$$\sqrt{\left(\frac{3va}{r}\right)^2 + \left(j - \frac{v^3}{r^2}\right)^2} \leq \hat{j}_{\max} \quad \forall d \in [-j_{\max}, j_{\max}] \quad (4.70)$$

$$\sqrt{\left(\frac{3a^2}{r} + \frac{4vj}{r} - \frac{v^4}{r^3}\right)^2 + \left(d - \frac{6v^2a}{r^2}\right)^2} \leq \hat{d}_{\max} \quad \forall d \in [-d_{\max}, d_{\max}] \quad (4.71)$$

Wie bei der Strecke gilt auch hier, dass der Wertebereich der kinematischen Bahngrößen durch die jeweiligen Restriktionen beschränkt ist. Somit kann die Bestimmung der Bahnbeschränkungen durch Lösen dieses Ungleichungssystems erfolgen, wobei die zugrundeliegenden Bahngrößen aus dem zulässigen Intervall gewählt werden müssen.

Dritter Lösungsschritt: Skalierung der Schranken

Da die Achsrestriktionen ein bestimmtes Verhältnis der einzelnen kinematischen Größen zueinander vorgeben und demzufolge ein entsprechendes Dynamikverhalten erwarten lassen, wäre es wünschenswert, dieses Verhalten auch auf die Bahnrestriktionen übertragen zu können. Zu diesem Zweck kann ein Skalierungsfaktor x eingeführt werden, der eine gleichzeitige Skalierung aller Bahnrestriktionen gewährleistet. Formal kann dieser Zusammenhang beschrieben werden, indem die maximal zulässigen Bahngrößen durch die skalierten oberen Schranken der Vektorlängen ausgedrückt werden:

$$v_{max} = x \cdot \hat{v}_{max} \quad (4.72)$$

$$a_{max} = x \cdot \hat{a}_{max} \quad (4.73)$$

$$j_{max} = x \cdot \hat{j}_{max} \quad (4.74)$$

$$d_{max} = x \cdot \hat{d}_{max} \quad (4.75)$$

$$\forall s \in [0, s_{ziel}]$$

$$\forall v \in [-x \cdot \hat{v}_{max}, x \cdot \hat{v}_{max}]$$

$$\forall a \in [-x \cdot \hat{a}_{max}, x \cdot \hat{a}_{max}] \quad \text{mit } x \in \mathbb{R}^+ \quad (4.76)$$

$$\forall j \in [-x \cdot \hat{j}_{max}, x \cdot \hat{j}_{max}]$$

$$\forall d \in [-x \cdot \hat{d}_{max}, x \cdot \hat{d}_{max}]$$

Die Zielstellung der folgenden Ausführungen besteht nun nicht mehr darin, die eigentlichen Bahnrestriktionen zu bestimmen, sondern es muss an deren Stelle lediglich der Skalierungsfaktor ermittelt werden. Auf dieser Basis kann die Untersuchung von Ungleichung (4.68) stattfinden. Im Sinne einer Worst-Case-Betrachtung kann die linke Seite der Ungleichung maximiert werden, indem v durch den größtmöglichen zulässigen Wert $x \cdot \hat{v}_{max}$ ersetzt wird. Da sowohl x als auch v_{max} positive reelle Werte darstellen, kann diese Gleichung problemlos nach x umgestellt werden:

$$x \leq 1 \quad (4.77)$$

In der gleichen Art und Weise kann nun auch die Ungleichung (4.69) untersucht werden. Nach Einsetzen der Worst-Case-Werte $v = x \cdot \hat{v}_{max}$ sowie $a = x \cdot \hat{a}_{max}$ und Umstellen nach x entsteht die folgende Ungleichung:

$$x \leq \sqrt{\frac{r \hat{a}_{max} \cdot \left(\sqrt{r^2 \hat{a}_{max}^2 + 4 \hat{v}_{max}^4} - r \hat{a}_{max} \right)}{2 \hat{v}_{max}^4}} \quad (4.78)$$

Der kleinere Wert für x aus den Gleichungen (4.77) und (4.78) begrenzt somit den zulässigen Wertebereich für x nach oben.

Die beiden zur Untersuchung verbleibenden Ungleichungen (4.70) und (4.71) können aufgrund ihrer Gestalt nur numerisch gelöst werden.

Vierter Lösungsschritt: Formulierung der Optimierungsprobleme

Im Folgenden sollen der Übersichtlichkeit halber lediglich die Optimierungsprobleme (OP) zur Lösung der Ungleichungen (4.70) und (4.71) formuliert werden. Ungleichung (4.70) kann gelöst werden durch:

OP 1: $\min_{v, a, j} |v|$, sodass

$$\sqrt{\left(\frac{3va}{r} \right)^2 + \left(j - \frac{v^3}{r^2} \right)^2} - \hat{j}_{max} = 0$$

$$a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$-j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

OP 2: $\min_{v, a, j} |a|$, sodass

$$\sqrt{\left(\frac{3va}{r}\right)^2 + \left(j - \frac{v^3}{r^2}\right)^2} - \hat{j}_{max} = 0$$

$$v - \frac{\hat{v}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

$$-v - \frac{\hat{v}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

$$j - \frac{\hat{j}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

$$-j - \frac{\hat{j}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

OP 3: $\min_{v, a, j} |j|$, sodass

$$\sqrt{\left(\frac{3va}{r}\right)^2 + \left(j - \frac{v^3}{r^2}\right)^2} - \hat{j}_{max} = 0$$

$$v - \frac{\hat{v}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$-v - \frac{\hat{v}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$a - \frac{\hat{a}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

Die anschließende Division der optimalen Zielfunktionswerte durch die entsprechenden Schranken liefert drei möglicherweise unterschiedliche Skalierungsfaktoren, von denen der kleinste die Lösung darstellt:

$$x \leq \min \left(\frac{|v|_{opt}}{\hat{v}_{max}}, \frac{|a|_{opt}}{\hat{a}_{max}}, \frac{|j|_{opt}}{\hat{j}_{max}} \right) \quad (4.79)$$

Auf dem gleichen Weg kann auch Ungleichung (4.71) gelöst werden. Da nun alle kinematischen Größen Bestandteil der Ungleichung sind, müssen folglich auch vier Optimierungsschritte durchgeführt werden. Die Problemstellung kann folgendermaßen formuliert werden:

OP 1: $\min_{v, a, j, d} |v|$, sodass

$$\begin{aligned} & \sqrt{\left(\frac{3a^2}{r} + \frac{4vj}{r} - \frac{v^4}{r^3} \right)^2 + \left(d - \frac{6v^2a}{r^2} \right)^2} - \hat{d}_{max} = 0 \\ & a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0 \\ & -a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0 \\ & j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0 \\ & -j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0 \\ & d - \frac{\hat{d}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0 \\ & -d - \frac{\hat{d}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0 \end{aligned}$$

OP 2: $\min_{v, a, j, d} |a|$, sodass

$$\begin{aligned} & \sqrt{\left(\frac{3a^2}{r} + \frac{4vj}{r} - \frac{v^4}{r^3} \right)^2 + \left(d - \frac{6v^2a}{r^2} \right)^2} - \hat{d}_{max} = 0 \\ & v - \frac{\hat{v}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0 \\ & -v - \frac{\hat{v}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0 \\ & j - \frac{\hat{j}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0 \\ & -j - \frac{\hat{j}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0 \\ & d - \frac{\hat{d}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0 \\ & -d - \frac{\hat{d}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0 \end{aligned}$$

OP 3: $\min_{v, a, j, d} |j|$, sodass

$$\sqrt{\left(\frac{3a^2}{r} + \frac{4vj}{r} - \frac{v^4}{r^3}\right)^2 + \left(d - \frac{6v^2a}{r^2}\right)^2} - \hat{d}_{max} = 0$$

$$v - \frac{\hat{v}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$-v - \frac{\hat{v}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$a - \frac{\hat{a}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$d - \frac{\hat{d}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$-d - \frac{\hat{d}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

OP 4: $\min_{v, a, j, d} |d|$, sodass

$$\sqrt{\left(\frac{3a^2}{r} + \frac{4vj}{r} - \frac{v^4}{r^3}\right)^2 + \left(d - \frac{6v^2a}{r^2}\right)^2} - \hat{d}_{max} = 0$$

$$v - \frac{\hat{v}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$-v - \frac{\hat{v}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$a - \frac{\hat{a}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$j - \frac{\hat{j}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$-j - \frac{\hat{j}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

Auf Basis der so bestimmten optimalen Werte kann auch hier die zulässige Skalierung mittels Auswahl des kleinsten Quotienten ermittelt werden:

$$x \leq \min \left(\frac{|v|_{opt}}{\hat{v}_{max}}, \frac{|a|_{opt}}{\hat{a}_{max}}, \frac{|j|_{opt}}{\hat{j}_{max}}, \frac{|d|_{opt}}{\hat{d}_{max}} \right) \quad (4.80)$$

An dieser Stelle liegen nun vier obere Schranken für den Skalierungsfaktor x vor. Die globale Obergrenze für x ist folglich der kleinste dieser vier Werte. Durch Einsetzen des Skalierungsfaktors in die Gleichungen (4.72) bis (4.75) können anschließend die gültigen Bahnrestriktionen ermittelt werden.

4.1.4 Kinematikplanung auf Bahnebene

Im vorangegangenen Abschnitt wurde erläutert, wie die kinematischen Restriktionen der Maschinenachsen auf Bahnebene projiziert werden können. Darüber hinaus wurde in Abschnitt 4.1.2.1 der geometrische Zusammenhang zwischen Trajektorie und Bahn gezeigt, was eine Transformation der Kinematik des Bahnparameters s in den \mathbb{R}^3 ermöglicht. Auf dieser Grundlage wird im Folgenden, nach einer kurzen Einführung in den Stand der Technik, die eindimensionale Kinematikplanung des Bahnparameters vorgestellt. Im Allgemeinen besteht das Ziel dieser Planungskomponente darin, in einer Dimension und unter Einhaltung vorgegebener Restriktionen eine festgelegte Bahnstrecke möglichst zeitoptimal zurückzulegen. Des Weiteren ist es Aufgabe der Bahnplanung, auch Qualitätsanforderungen an die Kinematikfunktionen, etwa bezüglich deren Stetigkeit und Differenzierbarkeit, zu berücksichtigen.

4.1.4.1 Stand der Technik auf dem Gebiet der analytischen Bahnplanung

In den letzten Jahrzehnten wurde eine Reihe von Bahnplanungsalgorithmen dritter und vierter Ordnung veröffentlicht, die geschmeidige Sollwerttrajektorien unter Berücksichtigung von kinematischen Beschränkungen erzeugen. Zum ersten Mal wurde eine analytische Trajektorienplanung dritter Ordnung von Olomski *et al.* veröffentlicht. Das kinematische Verhalten des Systems wird bei diesem Ansatz von einer stückweise konstanten Ruckfunktion bestimmt [79, 76, 89]. Die Trajektorien für Beschleunigung, Geschwindigkeit und Position ergeben sich anschließend durch mehrmalige Integration dieser Ruckfunktion. In Abbildung 4.5 sind die so erzeugten kinematischen Trajektorien beispielhaft angegeben. Es ist zu erkennen, dass das Ruckprofil ein „Bang Bang“ Verhalten aufweist und somit die geplanten Trajektorien zeitoptimalen Charakter besitzen.

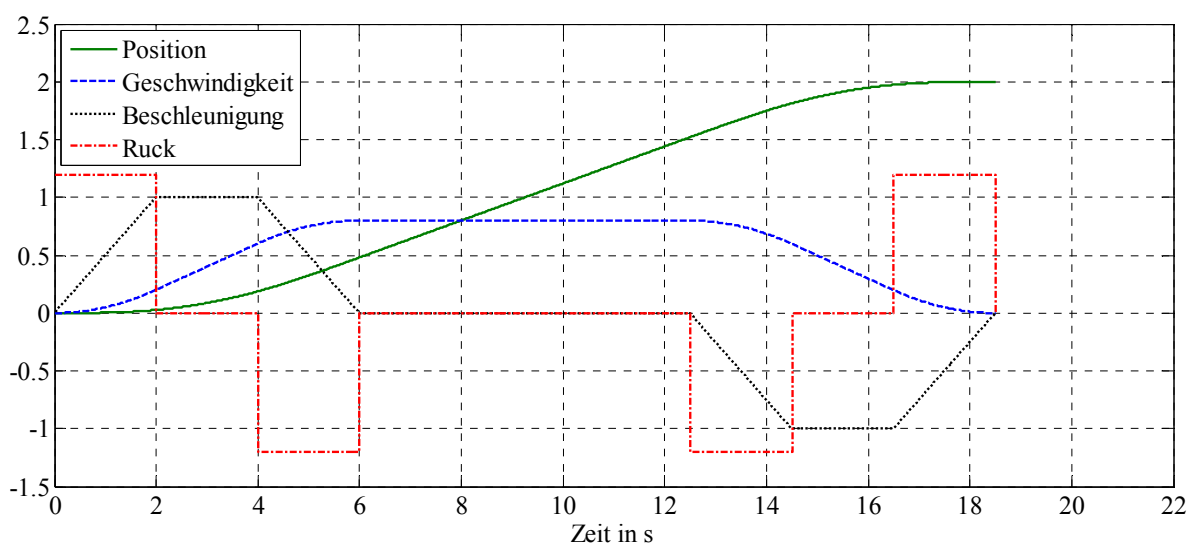


Abbildung 4.5: Beispiel für eine Bahnplanung 3. Ordnung mit einem stückweise konstanten Ruckprofil

Auf einen Beweis hierfür wird an dieser Stelle verzichtet und auf [82] verwiesen. Des Weiteren bietet dieser Algorithmus den Vorteil, dass er ausgesprochen recheneffizient arbeitet. Nachteilig wirkt sich jedoch die Tatsache aus, dass das geplante Ruckprofil einen nicht stetigen Verlauf aufweist und ein solches Verhalten niemals technisch realisiert werden kann.

Um dieses Problem zu beheben, stellte Sawodny *et al.* eine Methode vor, die das Ruckprofil aus kubischen Funktionen aufbaut [106, 107, 108] (siehe Abbildung 4.6). Alle von dieser Methode erzeugten Trajektorien sind technisch umsetzbar und so kommt es zu keiner ungewollten Anregung mechanischer Resonanzen. Der Vergleich von Abbildung 4.5 und Abbildung 4.6 zeigt jedoch, dass die mit diesem Algorithmus geplante Bewegung mehr Zeit benötigt als die identische Bewegung aus Abbildung 4.5. Die geringere Dynamik ist eine direkte Folge der kubischen Ruckfunktionen, denn durch diese können keine Phasen konstanter Beschleunigung realisiert werden. Darüber hinaus ist der rechentechnische Aufwand zur Parametrierung des kubischen Ruckprofiles erheblich höher als bei der Methode von Olomski *et al.*

Zur Vervollständigung sei an dieser Stelle noch der Algorithmus von Li *et al.* erwähnt, welcher anstelle von kubischen trigonometrische Funktionen zur Erzeugung des Ruckprofils nutzt [81]. Die Eigenschaften dieser Methode entsprechen der des Algorithmus von Sawodny *et al.*, und daher soll auf eine ausführliche Beschreibung verzichtet werden.

Die natürliche Erweiterung der bisher beschriebenen Algorithmen ist eine Bahnplanung vierter Ordnung. Mit dieser ist es möglich, zum einen stetige Ruckfunktionen zu erzeugen und zum anderen eine zeitoptimale Bewegung zu planen. Lambrechts *et al.* stellte hierzu eine Methode vor, die ähnlich zu Olomski *et al.* auf stückweise linearen Ruckanstiegsfunktionen basiert [70, 71]. Die Trajektorien für Ruck, Beschleunigung, Geschwindigkeit und Position ergeben sich auch hier zwangsläufig nach mehrmaliger Integration der Ruckanstiegsfunktion. Zur Veranschaulichung wurde die Bewegung aus Abbildung 4.5 und Abbildung 4.6 auch mit dieser Methode geplant. Abbildung 4.7 zeigt deutlich, dass die Dynamik der Bewegung nur unwesentlich geringer ist als bei Verwendung der Methode von Olomski *et al.* (siehe Abbildung 4.5).

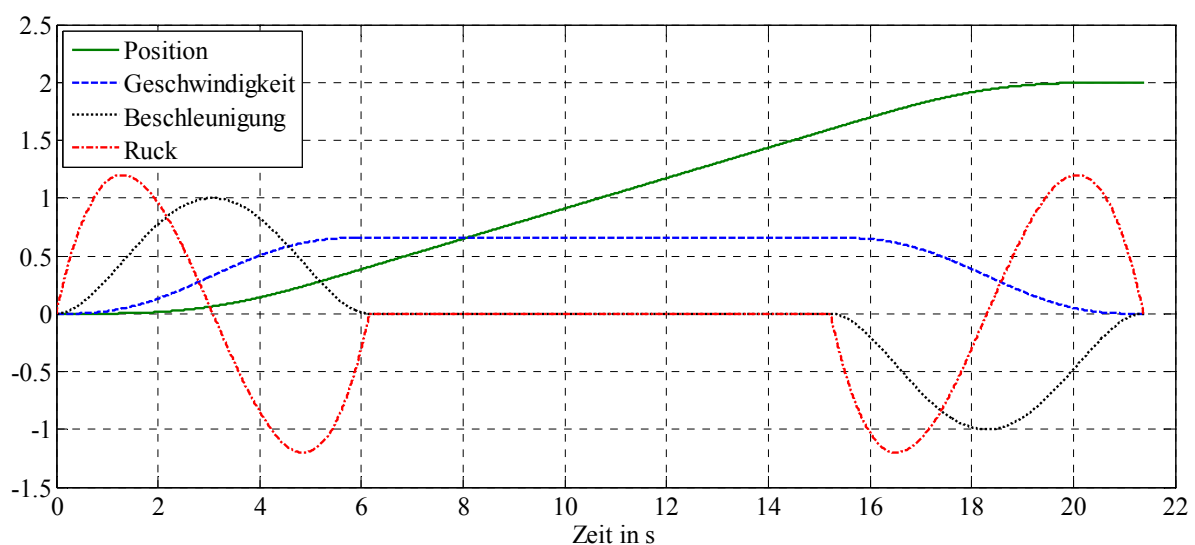


Abbildung 4.6: Beispiel für eine Bahnplanung 3.Ordnung mit einem stückweise kubischen Ruckprofil

Dieses Verhalten resultiert aus der Beschränkung des Ruckanstieges und somit kann zusammenfassend festgestellt werden, dass eine Bahnplanung vierter Ordnung es ermöglicht, die Vorteile der Methoden von Olomski *et al.* und Sawodny *et al.* zu vereinigen.

Ein entscheidender Nachteil aller vorgestellten Algorithmen ist jedoch, dass sie lediglich in der Lage sind, symmetrische Geschwindigkeitsprofile zu planen. Dieses nicht unerhebliche Defizit hat zur Folge, dass keine Bewegungen mit von Null verschiedenen Start- und/oder Endgeschwindigkeiten realisierbar sind. Daher ist es nicht möglich, mehrere verkettete Wegsegmente kontinuierlich zu durchfahren und dabei alle kinematischen Beschränkungen explizit zu berücksichtigen.

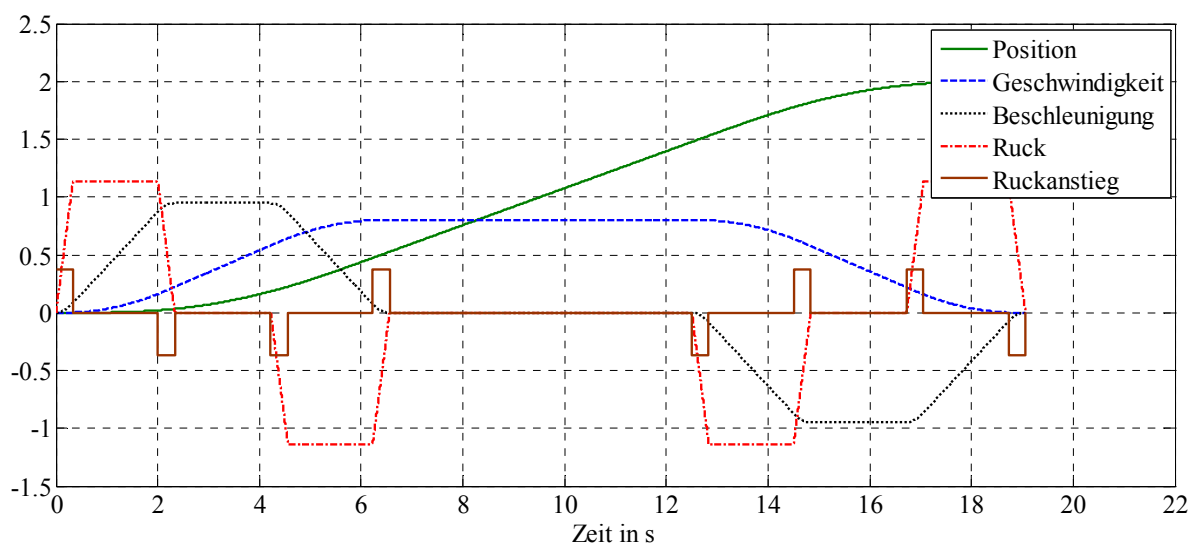


Abbildung 4.7: Beispiel für eine Bahnplanung 4. Ordnung mit einem stückweise konstanten Ruckanstieg

4.1.4.2 Bahnplanung vierter Ordnung

Zur Behebung der erwähnten Einschränkung wurde im Rahmen der vorliegenden Arbeit eine Bahnplanungsmethode vierter Ordnung entwickelt. Ziel des im Folgenden vorgestellten Verfahrens ist es, ein sequentielles Maximieren der kinematischen Größen nach dem Vorbild der Konstant-Ruck-Methode zu ermöglichen. Darüber hinaus sollte ein Algorithmus entwickelt werden, der es erlaubt, die Gesamtdynamik der Positionierung über mehrere Wegsegmente hinweg zu steigern. Dies ist nur realisierbar, wenn das kinematische System nicht an jedem Übergangspunkt zwischen den einzelnen Wegteilen zum Stillstand kommen muss. Dieses Ziel kann erreicht werden, indem die Übergänge verschliffen werden, was heißt, dass Richtungsänderungen langsam und kontinuierlich stattfinden und so mit relativ hoher Geschwindigkeit durchfahren werden können. Im Detail wird das Verschleifen der Basiselemente in Abschnitt 4.1.5 näher erläutert.

Zum Beginn der Herleitung stetig differenzierbarer Funktionen für die kinematischen Größen wurde in [131] folgende stetige Ruckanstiegsfunktion mitsamt der zugehörigen Restriktion neu eingeführt:

$$d(t) = \begin{cases} \frac{4d_{\max} t(T_1 - t)}{T_1^2} & \text{I1} \\ 0 & \text{I2} \\ \frac{4d_{\max} (T_1 + T_2 - t)(2T_1 + T_2 - t)}{T_1^2} & \text{I3} \\ 0 & \text{I4} \\ \frac{4d_{\max} (2T_1 + T_2 + T_3 - t)(3T_1 + T_2 - t)}{T_1^2} & \text{I5} \\ 0 & \text{I6} \\ -\frac{4d_{\max} (3T_1 + 2T_2 + T_3 - t)(4T_1 + 2T_2 + T_3 - t)}{T_1^2} & \text{I7} \\ 0 & \text{I8} \\ \frac{4d_{\max} (4T_1 + 2T_2 + T_3 + T_4 - t)(4T_1 + 2T_2 + T_3 + T_4 + T_5 - t)}{T_5^2} & \text{I9} \\ 0 & \text{I10} \\ -\frac{4d_{\max} (4T_1 + 2T_2 + T_3 + T_4 + T_5 + T_6 - t)(4T_1 + 2T_2 + T_3 + T_4 + 2T_5 + T_6 - t)}{T_5^2} & \text{I11} \\ 0 & \text{I12} \\ -\frac{4d_{\max} (4T_1 + 2T_2 + T_3 + T_4 + 2T_5 + T_6 + T_7 - t)(4T_1 + 2T_2 + T_3 + T_4 + 3T_5 + T_6 + T_7 - t)}{T_5^2} & \text{I13} \\ 0 & \text{I14} \\ \frac{4d_{\max} (4T_1 + 2T_2 + T_3 + T_4 + 3T_5 + 2T_6 + T_7 - t)(4T_1 + 2T_2 + T_3 + T_4 + 4T_5 + 2T_6 + T_7 - t)}{T_5^2} & \text{I15} \end{cases} \quad (4.81)$$

Nach mehrmaliger Integration ergeben sich daraus stetig differenzierbare Funktionsverläufe aller kinematischen Größen. Aus Gründen der Übersichtlichkeit wurden die Intervallgrenzen dieser stückweise definierten Ruckanstiegsfunktion verkürzt dargestellt. In Tabelle 4.1 werden den Intervallbezeichnern *I1* bis *I15* die jeweiligen Schranken zugeordnet.

Anhand der Berechnungsvorschrift (4.81) und der Tabelle 4.1 ist eine Unterteilung der Funktion in fünfzehn Teilbereiche erkennbar. Die Konstanten T_1 bis T_7 stehen dabei symbolisch für die sieben auftretenden Intervallbreiten, aus denen sich die tatsächliche Segmentierung des Definitionsbereichs ergibt (siehe Abbildung 4.8). Diese sieben Werte können a priori für jedes Wegsegment bestimmt werden und bleiben danach unveränderlich. Die Berechnung von T_1 bis T_7 geschieht auf Basis der gegebenen Restriktionen und der Weglänge. Bei näherer Untersuchung der Gleichung (4.81) fällt auf, dass die Teilfunktionen in den ungerade nummerierten Intervallen Polynome zweiten Grades darstellen. Die resultierenden Parabeln unterscheiden sich in ihrem Scheitelpunkt und in ihrer Öffnungsrichtung. Darüber hinaus variiert auch die Parabelbreite zwischen Beschleunigungs- und Bremsphase. Die genannten Parameter wurden derart gewählt, dass der Scheitelpunkt in Richtung der Abszisse genau in der jeweiligen Intervallmitte liegt und der zugehörige Funktionswert in seinem Betrag genau der Ruckanstiegsrestriktion d_{\max} entspricht.

Tabelle 4.1: Grenzen der Intervalle *I1* bis *I15*

Intervall	Intervallgrenzen
I1	$0 \leq t \leq T_1$
I2	$T_1 \leq t \leq T_1 + T_2$
I3	$T_1 + T_2 \leq t \leq 2T_1 + T_2$
I4	$2T_1 + T_2 \leq t \leq 2T_1 + T_2 + T_3$
I5	$2T_1 + T_2 + T_3 \leq t \leq 3T_1 + T_2 + T_3$
I6	$3T_1 + T_2 + T_3 \leq t \leq 3T_1 + 2T_2 + T_3$
I7	$3T_1 + 2T_2 + T_3 \leq t \leq 4T_1 + 2T_2 + T_3$
I8	$4T_1 + 2T_2 + T_3 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4$
I9	$4T_1 + 2T_2 + T_3 + T_4 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4 + T_5$
I10	$4T_1 + 2T_2 + T_3 + T_4 + T_5 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4 + T_5 + T_6$
I11	$4T_1 + 2T_2 + T_3 + T_4 + T_5 + T_6 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4 + 2T_5 + T_6$
I12	$4T_1 + 2T_2 + T_3 + T_4 + 2T_5 + T_6 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4 + 2T_5 + T_6 + T_7$
I13	$4T_1 + 2T_2 + T_3 + T_4 + 2T_5 + T_6 + T_7 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4 + 3T_5 + T_6 + T_7$
I14	$4T_1 + 2T_2 + T_3 + T_4 + 3T_5 + T_6 + T_7 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4 + 3T_5 + 2T_6 + T_7$
I15	$4T_1 + 2T_2 + T_3 + T_4 + 3T_5 + 2T_6 + T_7 \leq t \leq 4T_1 + 2T_2 + T_3 + T_4 + 4T_5 + 2T_6 + T_7$

Somit sind pro Parabel zwei Nullstellen vorhanden. Durch günstige Wahl der Parabelbreite fallen diese Nullstellen exakt auf die Intervallgrenzen und bilden so einen stetigen Übergang zu den Nullfunktionen in den geradzahlig benannten Intervallen. Beispielhaft werden diese Zusammenhänge in Abbildung 4.8 gezeigt. Wie zu sehen ist, wurde hier für die Ruckanstiegsrestriktion der Wert $d_{max} = 10 \text{ m/s}^4$ gewählt. Des Weiteren wird durch die sichtbare Trennung von Beschleunigungs- und Bremsphase bereits in der Ruckanstiegsfunktion ein Verhalten der Kinematik vorbereitet, das unterschiedliche Start- und Endgeschwindigkeiten zulässt. Aus der Ruckanstiegsfunktion können nun die weiteren kinematischen Größen auf Bahnebene ermittelt werden. Durch einmalige Integration von $d(t)$ über die Zeit t und anschließende geeignete Festlegung der Integrationskonstanten ergibt sich die Berechnungsvorschrift für die Ruckfunktion $j(t)$. Die Eignung der Werte für die Integrationskonstanten wird dabei durch zwei Forderungen bestimmt. Einerseits besteht auch für den Ruck das Ziel der Stetigkeit über alle fünfzehn Intervalle und andererseits muss die Ruckfunktion am Beginn und am Ende jedes Wegsegments Null sein.

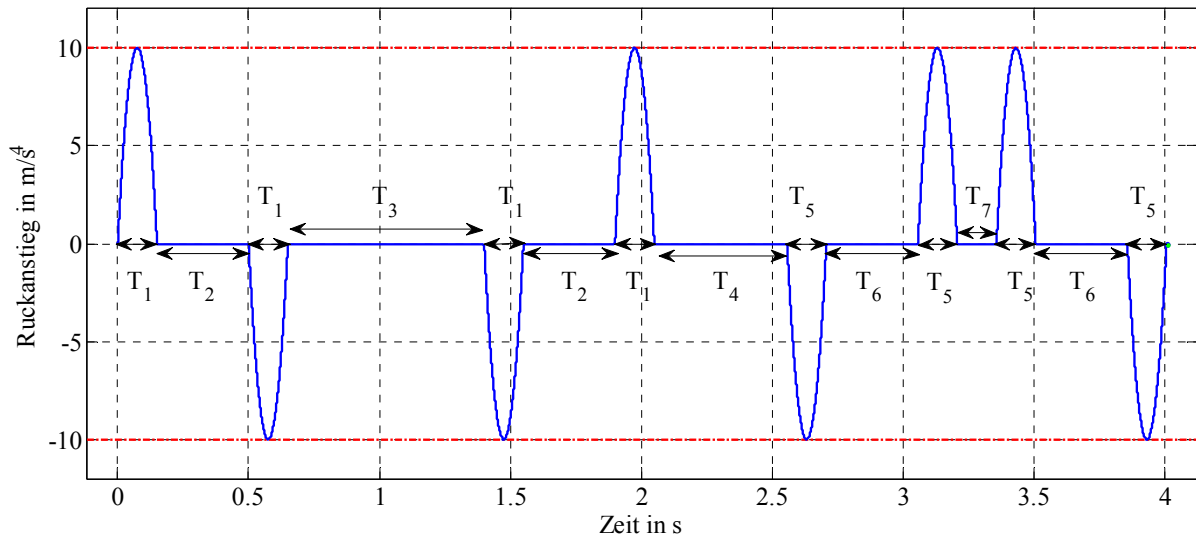


Abbildung 4.8: Beispiel einer Ruckanstiegsfunktion

Durch den integrativen Zusammenhang ist sichergestellt, dass die Ruckanstiegsfunktion exakt der Ableitung der Ruckfunktion entspricht. Da darüber hinaus beide Funktionen stetig gewählt wurden, ist die angestrebte stetige Differenzierbarkeit für den Ruck gegeben. Analog liefern weitere drei Integrationsschritte die stetig differenzierbaren Funktionen für die Beschleunigung, die Geschwindigkeit und den Ort. Auf die explizite Angabe der resultierenden Definitionen für $j(t)$, $a(t)$, $v(t)$ und $s(t)$ soll hier verzichtet werden, diese können im Anhang von [130] nachgeschlagen werden. Abbildung 4.9 zeigt die entsprechenden Verläufe, die unter Anwendung der hergeleiteten Funktionsdefinitionen entstehen. Die Restriktionen, die zur Erzeugung der Grafik herangezogen wurden, entsprechen dabei in ihren Werten denen aus Abbildung 4.8. Bei näherer Betrachtung von Abbildung 4.9 sind die geschilderten Zusammenhänge deutlich erkennbar. So entsprechen alle dort dargestellten kinematischen Funktionen den angestrebten Kriterien bezüglich Stetigkeit und stetiger Differenzierbarkeit. Aufgrund der Tatsache, dass die Beschleunigungs- und die Bremsphase getrennt voneinander parametrisiert werden können, entsteht ein „asymmetrischer“ Geschwindigkeitsverlauf.

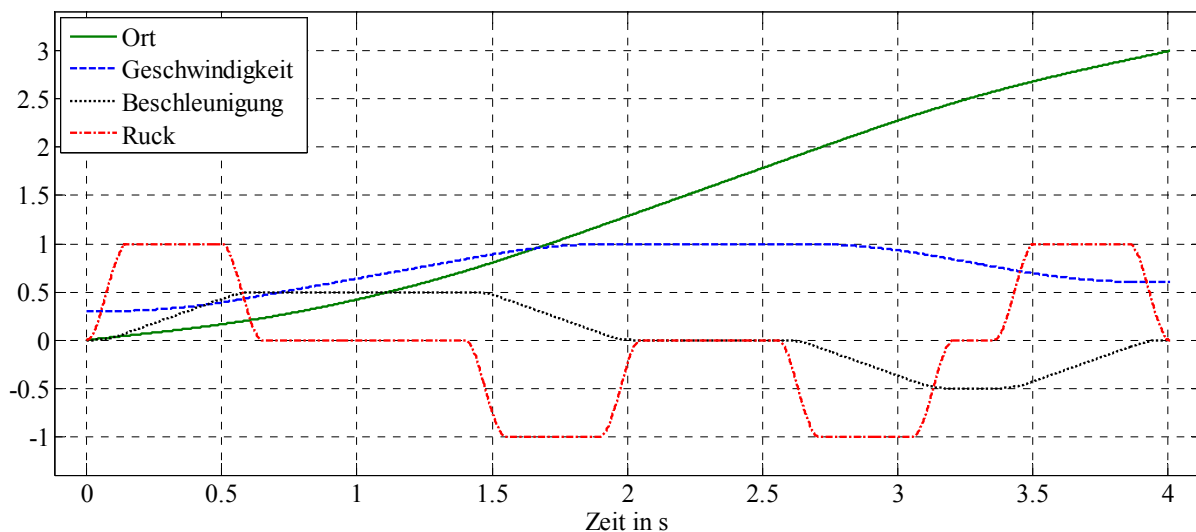


Abbildung 4.9: Beispiel für den Verlauf von Ruck, Beschleunigung, Geschwindigkeit und Ort

Der Begriff der Asymmetrie wurde in diesem Zusammenhang von [130] eingeführt und soll die Ungleichheit von Start- und Endgeschwindigkeit hervorheben. Im Beispiel aus Abbildung 4.9 wurden die Geschwindigkeiten an den Funktionsgrenzen zu $v_{start} = 0,3$ m/s und $v_{ziel} = 0,6$ m/s gewählt.

Kernkomponente des Planungsalgorithmus ist die Berechnung der Basisintervallbreiten T_1 bis T_7 , denn diese bestimmen direkt die Lage der Extrema der kinematischen Funktionen. Den Ausführungen in [130] folgend, können die Anforderungen Zeitoptimalität und Einhaltung der Beschränkungen nur erfüllt werden, indem eine kinematische Größe nach der anderen möglichst nahe an ihre Restriktion herangeführt wird. An dieser Stelle sei darauf hingewiesen, dass aus Sicht der Intervallbreitenbestimmung neben d_{max} , j_{max} , a_{max} , v_{max} auch s_{ziel} eine Restriktion darstellt. Zum Zweck der sequentiellen Ausnutzung der Beschränkungen wird in einem ersten Schritt die Intervallbreite der Ruckanstiegsphase während der Beschleunigung ermittelt. Begrenzend für den entsprechenden Wert T_1 wirkt die durch die resultierenden Kinematikfunktionen zuerst erreichte Restriktion. Dies kann mittels der folgenden Ungleichungen dargestellt werden:

$$j(t = T_1) \leq j_{max} \quad (4.82)$$

$$a(t = 2T_1 + T_2; T_2 = 0) \leq a_{max} \quad (4.83)$$

$$v(t = 4T_1 + 2T_2 + T_3; T_2, T_3 = 0) \leq v_{max} \quad (4.84)$$

$$s(t = 4T_1 + 2T_2 + T_3 + T_4 + 4T_5 + 2T_6 + T_7; T_2, T_3, T_4 = 0) \leq s_{ziel} \quad (4.85)$$

Durch Umstellen dieser Ungleichungen nach T_1 können theoretisch vier obere Schranken für die Intervallbreite der Ruckanstiegsphase bestimmt werden. Die Auswahl des kleinsten der berechneten Werte ergibt T_1 . Allerdings gelingt diese Umformung nicht für die Ungleichung (4.85), denn diese enthält neben T_1 auch die Unbekannten T_5 , T_6 und T_7 . Zur Lösung des Problems wird die Wegrestriktion im ersten Schritt der Intervallbreitenbestimmung vernachlässigt und Beschleunigungs- und Bremsphase werden losgelöst voneinander betrachtet. Die Berechnung erfolgt somit rein analytisch auf Basis von Gleichung (4.82) bis (4.84). Unter der Annahme, dass T_1 mit einem korrekten Wert versehen werden konnte, folgt im zweiten Schritt die Berechnung von T_2 in analoger Art und Weise. Die entsprechenden Ungleichungen lauten:

$$a(t = 2T_1 + T_2; T_1 \text{ bekannt}) \leq a_{max} \quad (4.86)$$

$$v(t = 4T_1 + 2T_2 + T_3; T_1 \text{ bekannt}; T_3 = 0) \leq v_{max} \quad (4.87)$$

$$s(t = 4T_1 + 2T_2 + T_3 + T_4 + 4T_5 + 2T_6 + T_7; T_1 \text{ bekannt}; T_3, T_4 = 0) \leq s_{ziel} \quad (4.88)$$

Auch hier kann die Lösung nur unter Vernachlässigung der Ortsrestriktion analytisch erfolgen. Nach diesem Vorbild kann die Berechnung der restlichen fünf Intervallbreiten fortgesetzt werden.

Im zweiten Schritt wird nun geprüft, ob die Wegrestriktion verletzt wird. Dies kann stattfinden, indem die Phase konstanter Geschwindigkeit wie folgt ermittelt wird:

$$\begin{aligned}
s(t = 4T_1 + 2T_2 + T_3 + T_4 + 4T_5 + 2T_6 + T_7; T_1, T_2, T_3, T_5, T_6, T_7 \text{ bekannt}) &\leq s_{\text{ziel}} \\
\Rightarrow T_4 &\leq -\frac{1}{6d_{\max} T_1^2 T_2 + 2d_{\max} T_1 T_2 T_3 + 4d_{\max} T_1^3 + 2d_{\max} T_1 T_2^2 + 2d_{\max} T_1^2 T_3} \cdot \\
&\quad \left(\begin{aligned}
&3d_{\max} T_3 T_1 T_2^2 - d_{\max} T_5 T_6 T_7^2 + 2d_{\max} T_1 T_2 T_3 T_7 + 4d_{\max} T_1 T_6 T_2^2 \\
&+ 24d_{\max} T_1^2 T_2 T_5 - d_{\max} T_1^2 T_3^3 - 3d_{\max} T_5 T_6^2 T_7 + 10d_{\max} T_1^2 T_2^2 \\
&+ 12d_{\max} T_1^2 T_2 T_6 + 9d_{\max} T_1^2 T_2 T_3 - 9d_{\max} T_5^2 T_6 T_7 + 2d_{\max} T_1 T_2^2 T_7 \\
&- d_{\max} T_5^2 T_7^2 + 16d_{\max} T_1^3 T_2 + 8d_{\max} T_1^2 T_3 T_5 + 4d_{\max} T_1^3 T_7 + 8d_{\max} T_1 T_2^2 T_5 \\
&+ 6d_{\max} T_1^2 T_2 T_7 + 6T_2 v_{\text{start}} + 3T_3 v_{\text{start}} 6T_6 v_{\text{start}} + 3T_7 v_{\text{start}} + 12T_1 v_{\text{start}} \\
&+ 6d_{\max} T_1^3 T_3 + 2d_{\max} T_1 T_2^3 + 4d_{\max} T_1^2 T_3 T_6 + 12T_5 v_{\text{start}} + 2d_{\max} T_1^2 T_3 \\
&- 8d_{\max} T_5^4 - 10d_{\max} T_5^2 T_6^3 + 8d_{\max} T_1 T_2 T_3 T_5 + 8d_{\max} T_1^3 T_6 - 2d_{\max} T_5 T_6^3 \\
&+ 4d_{\max} T_1 T_2 T_3 T_6 - 6d_{\max} T_5^3 T_7 - 2s_{\text{ziel}} + 8d_{\max} T_1^4 + d_{\max} T_1 T_2 T_3^2 \\
&- 16d_{\max} T_5^3 T_6 + 16d_{\max} T_1^3 T_5
\end{aligned} \right) \quad (4.89)
\end{aligned}$$

Ergibt sich ein negatives T_4 , so bedeutet dies, dass die berechnete Beschleunigungs- und Bremsphase mehr Weg benötigen als zur Verfügung steht. Somit sind alle Ergebnisse ungültig und die Intervallbreiten müssen numerisch bestimmt werden [131]. Da eine numerische Lösung jedoch mit einer ganzen Reihe von Nachteilen einhergeht, werden alle Spezialfälle, die eine direkte analytische Lösung ermöglichen, vorher vom Algorithmus abgefangen. Beispielsweise fällt die komplette Bremsphase weg, falls die Endgeschwindigkeit der maximal zulässigen Geschwindigkeit entspricht. Dadurch können die Parameter T_5 , T_6 und T_7 aus der Ortsfunktion eliminiert werden. Die resultierenden Ungleichungen können anschließend problemlos umgestellt und analytisch gelöst werden. Des Weiteren besteht unter der Randbedingung $v_{\text{ziel}} = v_{\text{max}}$ die Möglichkeit, die Werte für T_1 bis T_4 ohne Einsatz numerischer Verfahren zu bestimmen. Analog entfällt für $v_{\text{start}} = v_{\text{max}}$ die Beschleunigungsphase und T_4 bis T_7 sind problemlos berechenbar. Die Kombination beider Fälle bildet den dritten berücksichtigten Fall, in dem ausschließlich eine Phase konstanter Geschwindigkeit auftritt und somit einzig das korrespondierende T_4 bestimmt werden muss. Auch wurde in [130] gezeigt, dass der Spezialfall $v_{\text{start}} = v_{\text{ziel}} = 0$ aufgrund seiner Symmetrie und durch den Wegfall der beiden Randgeschwindigkeiten ebenfalls eine analytische Berechnung zulässt.

4.1.5 Kontinuierliche Bahnplanung mittels Verschleifelementen

Sollen mehrere Messfahrten hintereinander ausgeführt werden, ist es wünschenswert, diese Wegsegmente kontinuierlich und rucksprungfrei zu durchfahren. Die benutzerdefinierten Messstrecken werden durch die Basiselemente Strecke und Kreisbogen beschrieben. Aufgrund der unterschiedlichen Krümmung dieser Elemente ist eine kontinuierliche Durchfahrt der Gesamttrajektorie mit einem stetigen Beschleunigungsprofil nicht möglich. Zur Lösung des Problems wurden in [82] zusätzlich zu den bereits vorgestellten Basiselementen weitere Kurvenformen eingeführt. Hierbei handelt es sich um die Bloss- und die erweiterte Blosskurve. Die Blosskurve besitzt einen quadratischen Krümmungsanstieg und ist daher prädestiniert für den Einsatz als Verschleifelement. Der in dieser Arbeit entwickelte Algorithmus ist in der Lage, zwei sich schneidende Geraden zu verbinden (siehe Abbildung 4.10). Des Weiteren

können Tangente und Kreis ohne tangentiale Abrückung (siehe Abbildung A 14 in Anhang B), Tangente und Kreis mit tangentialer Abrückung (siehe Abbildung A 15) sowie Sekante und Kreis (siehe Abbildung A 16) verschliffen werden. Hierbei erfolgt je nach vorliegender Paarung die Verschleifung durch eine Kombination aus mindestens einer Blosskurve und, falls notwendig, einem Kreisbogen. Im Folgenden soll für die Bloss- und die erweiterte Blosskurve sowohl die Darstellung in Abhängigkeit des Bahnortes als auch die Restriktionsbestimmung auf Bahnebene hergeleitet werden. Da die Herleitungen eine sehr große Ähnlichkeit zu den Ausführungen am Beispiel des Kreisbogens aufweisen, kann diese Betrachtung sehr stark verkürzt stattfinden. Da die Blosskurve einen Spezialfall der erweiterten Blosskurve darstellt, beschränken sich die angegebenen Gleichungen auf die allgemeinere Form.

Die erweiterte Blosskurve wurde in [82] mit folgender Ortsfunktion eingeführt:

$$\tilde{\vec{s}} = \int_0^{s_{\text{ziel}}} \begin{pmatrix} \cos \left(\frac{s^3 (s - 2s_{\text{ziel}}) \left(\frac{1}{r_{\text{start}}} - \frac{1}{r_{\text{ziel}}} \right)}{2s_{\text{ziel}}^2} + \frac{1}{r_{\text{start}}} \right) \\ \sin \left(\frac{s^3 (s - 2s_{\text{ziel}}) \left(\frac{1}{r_{\text{start}}} - \frac{1}{r_{\text{ziel}}} \right)}{2s_{\text{ziel}}^2} + \frac{1}{r_{\text{start}}} \right) \\ 0 \end{pmatrix} ds \quad (4.90)$$

Es sei angemerkt, dass analog zum Kreisbogen in (4.90) bereits eine Vereinfachung mittels geeigneter Koordinatentransformationen durchgeführt wurde. Darüber hinaus geht die erweiterte Blosskurve in die Blosskurve über, wenn $r_{\text{start}} \rightarrow \infty$ bzw. $1/r_{\text{start}} = 0$ gilt.

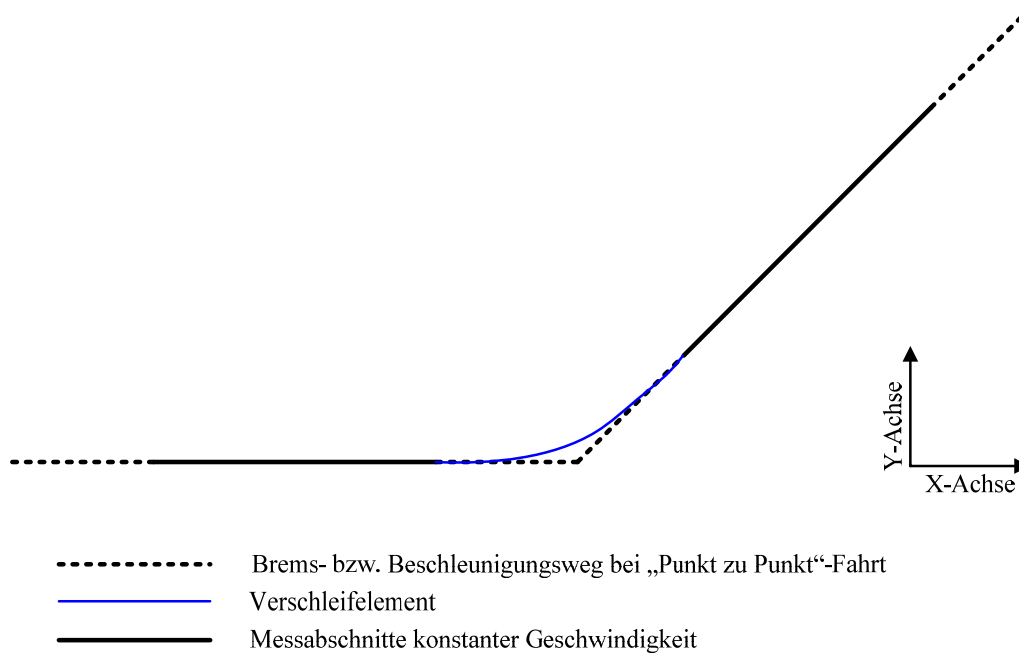


Abbildung 4.10: Beispiel für das Verschleifen zweier Strecken

Dies ist nachvollziehbar, da r_{start} den Radius des Schmiegekreises am Kurvenanfang darstellt, während r_{ziel} den entsprechenden Zusammenhang am Kurvenende festlegt. Im Falle der einfachen Blosskurve liegt zu Beginn keine Krümmung vor, was durch $r_{start} \rightarrow \infty$ zum Ausdruck kommt. Problematisch an Gleichung (4.90) ist jedoch, dass die Stammfunktion nicht analytisch berechnet werden kann, und somit ein echtzeitfähiges numerisches Lösungsverfahren zum Einsatz kommen muss. Ein Ansatz hierzu bietet die „Gauß-Legendrische“-Quadratur, welche die benötigte Approximationsgenauigkeit bei geringer Rechenlast bietet. Einen tieferen Einblick in die Herleitung sowie die Umsetzung zur Integration von Gleichung (4.90) gibt [82].

4.1.5.1 Bestimmung der Kurvenkinematik für die erweiterte Blosskurve

Mittels Differentiation kann aus (4.90) das Frenetsche Dreiein inklusive der zugehörigen Koeffizienten bestimmt werden:

$$\tilde{\mathbf{t}} = \begin{pmatrix} \cos \left(\frac{s^3(s-2s_{ziel}) \left(\frac{1}{r_{start}} - \frac{1}{r_{ziel}} \right)}{2s_{ziel}^2} + \frac{s}{r_{start}} \right) \\ \sin \left(\frac{s^3(s-2s_{ziel}) \left(\frac{1}{r_{start}} - \frac{1}{r_{ziel}} \right)}{2s_{ziel}^2} + \frac{s}{r_{start}} \right) \\ 0 \end{pmatrix} \quad (4.91)$$

$$\tilde{\mathbf{n}} = \begin{pmatrix} -\sin \left(\frac{s^3(s-2s_{ziel}) \left(\frac{1}{r_{start}} - \frac{1}{r_{ziel}} \right)}{2s_{ziel}^2} + \frac{s}{r_{start}} \right) \\ \cos \left(\frac{s^3(s-2s_{ziel}) \left(\frac{1}{r_{start}} - \frac{1}{r_{ziel}} \right)}{2s_{ziel}^2} + \frac{s}{r_{start}} \right) \\ 0 \end{pmatrix} \quad (4.92)$$

$$\tilde{\mathbf{b}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (4.93)$$

$$\kappa = \frac{s^2(2s-3s_{ziel}) \left(\frac{1}{r_{start}} - \frac{1}{r_{ziel}} \right)}{s_{ziel}^3} + \frac{1}{r_{start}} \quad (4.94)$$

$$\kappa' = \frac{6s(s - s_{\text{ziel}})(r_{\text{ziel}} - r_{\text{start}})}{r_{\text{start}}r_{\text{ziel}}s_{\text{ziel}}^3} \quad (4.95)$$

$$\kappa'' = \frac{6(2s - s_{\text{ziel}})(r_{\text{ziel}} - r_{\text{start}})}{r_{\text{start}}r_{\text{ziel}}s_{\text{ziel}}^3} \quad (4.96)$$

$$\tau = 0 \quad (4.97)$$

Die Tatsache, dass auch die Blosskurven ebene Trajektorien darstellen, führt zu einem konstanten Binormalenvektor und zu einem Wegfall der Torsion. Des Weiteren ist zu bemerken, dass κ im Gegensatz zum Kreisbogen keine konstante Funktion darstellt, und somit die ersten beiden Ableitungen von Null verschieden sind. Bei Anwendung der Gleichungen (4.8) bis (4.11) ergeben sich die folgenden Berechnungsvorschriften für die Kinematik der erweiterten Blosskurve im transformierten Koordinatensystem:

$$\tilde{\mathbf{v}} = \tilde{\mathbf{t}} \cdot \mathbf{v} \quad (4.98)$$

$$\tilde{\mathbf{a}} = \tilde{\mathbf{n}} \cdot \kappa \cdot v^2 + \tilde{\mathbf{t}} \cdot \mathbf{a} \quad (4.99)$$

$$\tilde{\mathbf{j}} = \tilde{\mathbf{n}} \cdot (\kappa' v^3 + 3\kappa v a) + \tilde{\mathbf{t}} \cdot (j - \kappa^2 v^3) \quad (4.100)$$

$$\tilde{\mathbf{d}} = \tilde{\mathbf{n}} \cdot (3\kappa a^2 + 4\kappa v j - \kappa^3 v^4 + 6\kappa' v^2 a + \kappa'' v^4) + \tilde{\mathbf{t}} \cdot (d - 6\kappa^2 v^2 a - 3 \cdot \kappa \kappa' v^4) \quad (4.101)$$

4.1.5.2 Bestimmung der Bahnrestriktionen für die erweiterte Blosskurve

Auch die Bestimmung der Bahnrestriktionen erfolgt unter Anwendung des Ansatzes, der bereits für den Kreisbogen skizziert wurde. Voraussetzung ist auch hier die erfolgreiche Berechnung der oberen Schranken für die Vektorlängen. Diese können, wie bereits bekannt, über den Schnitt der Kurvenebene mit den Koordinatenebenen ermittelt werden. Unter Einsatz des Satzes des Pythagoras können auf Basis der Gleichungen (4.98) bis (4.101) die folgenden Berechnungsvorschriften gebildet werden:

$$\sqrt{v^2} \leq \hat{v}_{\text{max}} \quad (4.102)$$

$$\sqrt{(\kappa v^2)^2 + a^2} \leq \hat{a}_{\text{max}} \quad (4.103)$$

$$\sqrt{(\kappa' v^3 + 3\kappa v a)^2 + (j - \kappa^2 v^3)^2} \leq \hat{j}_{\text{max}} \quad (4.104)$$

$$\sqrt{(3\kappa a^2 + 4\kappa v j - \kappa^3 v^4 + 6\kappa' v^2 a + \kappa'' v^4)^2 + (d - 6\kappa^2 v^2 a - 3 \cdot \kappa \kappa' v^4)^2} \leq \hat{d}_{\text{max}} \quad (4.105)$$

$$\forall s \in [0, s_{\text{ziel}}]$$

$$\forall v \in [-x \cdot \hat{v}_{\text{max}}, x \cdot \hat{v}_{\text{max}}]$$

$$\forall a \in [-x \cdot \hat{a}_{\text{max}}, x \cdot \hat{a}_{\text{max}}] \quad \text{mit } x \in \mathbb{R}^+ \quad (4.106)$$

$$\forall j \in [-x \cdot \hat{j}_{\text{max}}, x \cdot \hat{j}_{\text{max}}]$$

$$\forall d \in [-x \cdot \hat{d}_{\text{max}}, x \cdot \hat{d}_{\text{max}}]$$

Auf Grundlage der Vorschriften (4.102) bis (4.106) kann nun die Bestimmung des Skalierungsfaktors x stattfinden. Da Ungleichung (4.102) exakt mit (4.68) übereinstimmt, kann die gleiche Lösung wie für den Kreisbogen ermittelt werden:

$$x \leq 1 \quad (4.107)$$

Auch Ungleichung (4.103) kann erneut analytisch gelöst werden. Im Sinne einer Worst-Case-Betrachtung können die Variablen v und a durch $x \cdot \hat{v}_{\max}$ bzw. $x \cdot \hat{a}_{\max}$ ersetzt werden. Der größtmögliche Wert für κ ist anschließend durch eine Kurvendiskussion ermittelbar:

$$\kappa' = 0 \Leftrightarrow s = 0 \vee s = s_{\text{ziel}}$$

$$\kappa''(0) = \frac{6(r_{\text{start}} - r_{\text{ziel}})}{r_{\text{start}} r_{\text{ziel}} s_{\text{ziel}}^2}, \kappa''(s_{\text{ziel}}) = \frac{6(r_{\text{ziel}} - r_{\text{start}})}{r_{\text{start}} r_{\text{ziel}} s_{\text{ziel}}^2}$$

Fall 1 :

$$r_{\text{ziel}} > r_{\text{start}} > 0$$

$$\Rightarrow \kappa''(0) < 0, \kappa''(s_{\text{ziel}}) > 0$$

$$\Rightarrow \text{Maximum } \kappa(0) = \frac{1}{r_{\text{start}}}, \text{ Minimum } \kappa(s_{\text{ziel}}) = \frac{1}{r_{\text{ziel}}}$$

$$\Rightarrow \text{betragsgrößte Krümmung } \kappa = \frac{1}{r_{\text{start}}}$$

Fall 2 :

$$r_{\text{start}} > r_{\text{ziel}} > 0$$

$$\Rightarrow \kappa''(0) > 0, \kappa''(s_{\text{ziel}}) < 0$$

$$\Rightarrow \text{Minimum } \kappa(0) = \frac{1}{r_{\text{start}}}, \text{ Maximum } \kappa(s_{\text{ziel}}) = \frac{1}{r_{\text{ziel}}}$$

$$\Rightarrow \text{betragsgrößte Krümmung } \kappa = \frac{1}{r_{\text{ziel}}}$$

Weitere Fälle müssen nicht berücksichtigt werden, da negative und identische Radien ausgeschlossen werden. Auch die Grenzwerte $s \rightarrow \pm\infty$ sind irrelevant, da s durch (4.106) entsprechend beschränkt ist. Die dargestellte Herleitung zeigt, dass der Worst-Case-Wert für κ dem Kehrwert des kleineren der beiden Radien r_{start} und r_{ziel} entspricht. In beiden behandelten Fällen geht (4.103) in eine zu (4.69) äquivalente Form über und es gilt:

$$x \leq \sqrt{\frac{r_{\text{start}} \hat{a}_{\max} \cdot \left(\sqrt{r_{\text{start}}^2 \hat{a}_{\max}^2 + 4 \hat{v}_{\max}^4} - r_{\text{start}} \hat{a}_{\max} \right)}{2 \hat{v}_{\max}^4}} \quad \text{falls } r_{\text{start}} < r_{\text{ziel}} \quad (4.108)$$

$$x \leq \sqrt{\frac{r_{\text{ziel}} \hat{a}_{\max} \cdot \left(\sqrt{r_{\text{ziel}}^2 \hat{a}_{\max}^2 + 4 \hat{v}_{\max}^4} - r_{\text{ziel}} \hat{a}_{\max} \right)}{2 \hat{v}_{\max}^4}} \quad \text{falls } r_{\text{ziel}} < r_{\text{start}} \quad (4.109)$$

Somit können, wie beim Kreisbogen, die ersten beiden Ungleichungen analytisch gelöst werden. Für die verbleibenden beiden Ungleichungen (4.104) und (4.105) kann nun erneut das bereits beschriebene Optimierungsverfahren Anwendung finden (siehe Anhang B). Hierbei gilt es zu beachten, dass s in den Satz der Variablen aufgenommen werden muss, denn durch die Krümmungsfunktion tritt auch der Bahnweg innerhalb der Gleichungsnebenbedingungen auf. Die verbleibenden Berechnungsschritte, die zur Bestimmung der Skalierungen und damit der Bahnrestriktionen notwendig sind, entsprechen exakt denen des Kreisbogens. Auch die

technische Umsetzung zeigt keine wesentlichen Veränderungen über die angepassten Optimierungsschritte hinaus. Sind die Restriktionen auf Bahnebene bekannt, kann die Kinematikberechnung auf dem Verschleifelement mit dem in Abschnitt 4.1.4.2 vorgestellten Algorithmus durchgeführt werden.

4.1.5.3 Bestimmung der Übergangsgeschwindigkeiten

In den vorangegangenen Abschnitten wurden die Grundlagen der Bahnplanung hergeleitet. Des Weiteren wurde aufgezeigt, wie durch Verschleifung von mehreren Basiselementen eine Trajektorie erzeugt werden kann. Der Bahnplaner erzeugt nun für jedes einzelne Element die kinematischen Verläufe auf Bahnebene (siehe Abschnitt 4.1.4). Hierbei sind alle höheren zeitlichen Ableitungen der Geschwindigkeit am Elementanfang und -ende Null und müssen daher entlang der Gesamtstrecke nicht weiter beachtet werden. Für die Bahngeschwindigkeit gilt dies jedoch nicht, denn diese soll über die Gesamtstrecke einen kontinuierlichen Verlauf besitzen. In diesem Abschnitt wird eine Methode vorgestellt, mit welcher diese Start- und Endgeschwindigkeiten bestimmt werden können. Somit wird es dem Bahnplanungsalgorithmus ermöglicht, die Kinematik elementweise zu berechnen, ohne dabei die Gesamtstrecke betrachten zu müssen. Auf Bahnebene ist ein Element durch die sieben Größen s_{ziel} , v_{max} , a_{max} , j_{max} , d_{max} , v_{start} und v_{ziel} gekennzeichnet. Diese Größen des i -ten Streckenelementes können zu einem Elementvektor

$$e_i = \begin{pmatrix} s_{ziel}^{(i)} \\ v_{max}^{(i)} \\ a_{max}^{(i)} \\ j_{max}^{(i)} \\ d_{max}^{(i)} \\ v_{start}^{(i)} \\ v_{ziel}^{(i)} \end{pmatrix} \quad (4.110)$$

zusammengefasst werden, welcher das entsprechende Streckenelement vollständig beschreibt. Da nun der Bahnplaner zu jedem Streckenelement e_i gesondert die Bahnkinematiken berechnet, wird ein Algorithmus benötigt, welcher alle $v_{start}^{(i)}$ und $v_{ziel}^{(i)}$ unter Beachtung der gesamten Trajektorie ermittelt. Der Lösungsansatz besteht darin, die Geschwindigkeiten unter den gegebenen Restriktionen immer maximal zu wählen. Dafür ist es notwendig, ausgehend von der Startgeschwindigkeit v_{start} , die unter den gegebenen Beschränkungen maximal mögliche Endgeschwindigkeit v_{ziel} zu ermitteln. In Anlehnung an die von dem Bahnplanungsalgorithmus genutzten Beziehungen für die Berechnung der Bahnkinematik kann das zu lösende Problem wie folgt allgemein formuliert werden:

$$v_{ziel} = f(s_{ziel}, v_{max}, a_{max}, j_{max}, d_{max}, v_{start}) \quad (4.111)$$

Hiermit wird es möglich, in Abhängigkeit der in (4.111) angegebenen Größen die maximal erreichbare Endgeschwindigkeit auf einem Streckenelement zu berechnen. Es können prinzi-

piell drei mögliche Fälle der Start- und Endgeschwindigkeiten in Bezug zur maximalen Bahngeschwindigkeit v_{max} auftreten (siehe Abbildung 4.11). Start- und Endgeschwindigkeit können kleiner v_{max} sein, welches dem ersten Fall in Abbildung 4.11 entspricht. In Fall 2 ist die Startgeschwindigkeit kleiner und die Endgeschwindigkeit größer oder gleich der Bahngeschwindigkeitsbegrenzung. Die verbleibende Variante ist in Fall 3 dargestellt, in dem v_{start} größer oder gleich v_{max} und v_{ziel} größer v_{max} ist. Tritt Fall 1 ein, so wird die berechnete Endgeschwindigkeit übernommen und die aktuelle Startgeschwindigkeit nicht modifiziert. Weiterhin wird zunächst die Endgeschwindigkeit $v_{ziel}^{(i)}$ als Startgeschwindigkeit des nachfolgenden Elementes $v_{start}^{(i+1)}$ angenommen, sodass gilt:

$$v_{start}^{(i+1)} = v_{ziel}^{(i)} \quad (4.112)$$

Gilt im zweiten Fall $v_{ziel} > v_{max}$, so wird

$$v_{max}^{(i)} = v_{ziel}^{(i)} = v_{start}^{(i+1)} \quad (4.113)$$

gesetzt. Die Startgeschwindigkeit des aktuellen Elementes bleibt hierbei unverändert. Im letzten Fall wird die Startgeschwindigkeit zu

$$v_{max}^{(i)} = v_{start}^{(i)} = v_{ziel}^{(i-1)} \quad (4.114)$$

festgelegt. Darüber hinaus muss die Endgeschwindigkeit des vorherigen Streckenelementes begrenzt werden, damit der Verlauf der Bahngeschwindigkeiten entlang der gesamten Wegstrecke stetig ist.

Die Berechnung der Übergangsgeschwindigkeiten soll nun anhand eines Beispiels kurz skizziert werden. Die Startgeschwindigkeit des ersten Streckenelementes ist zwingend $v_{start}^{(1)} = 0$. Somit kann für dieses Element die Endgeschwindigkeit mithilfe des Bahnplanungsalgorithmus ermittelt werden, wobei hier lediglich der Fall 1 oder 2 auftreten kann. Die Startgeschwindigkeit des nachfolgenden Streckenelementes wird zunächst zu $v_{start}^{(2)} = v_{ziel}^{(1)}$ bei Fall 1 und zu $v_{start}^{(2)} = v_{max}^{(1)}$ bei Fall 2 angenommen. In jedem weiteren Element folgt als Erstes die Prüfung der Startgeschwindigkeit im Bezug auf die maximale Elementgeschwindigkeit, um Fall 3 auszuschließen oder zu bestätigen. In diesem Beispiel soll angenommen werden, dass auch im zweiten Streckenelement Fall 1 oder 2 zutrifft. Bei der Betrachtung des dritten Streckenelementes wird wieder die Startgeschwindigkeit ausgewertet. Hier soll nun $v_{start}^{(3)} > v_{max}^{(3)}$ gelten und somit der dritte Fall vorliegen. Demnach wird $v_{ziel_{neu}}^{(2)} = v_{start}^{(3)} = v_{max}^{(3)}$ gesetzt. Im Falle dass $v_{ziel_{neu}}^{(2)}$ kleiner ist als $v_{start}^{(2)}$, entsteht die Fragestellung, ob der Weg des zweiten Elementes ausreicht, um von $v_{start}^{(2)}$ auf $v_{ziel_{neu}}^{(2)}$ abzubremesen. Sollte dies nicht der Fall sein, so müsste auch die Startgeschwindigkeit des zweiten Elementes $v_{start}^{(2)}$ entsprechend angepasst werden. Somit ist klar zu erkennen, dass sich bei ungünstigen Randbedingungen die Veränderung einer Start- oder Endgeschwindigkeit über alle Basiselemente der Trajektorie hinweg fortpflanzen kann. Diese rückwärtige Ausbreitung erstreckt sich über alle n zusammenhängenden Elemente für

die $v_{ziel}^{(j)} < v_{start}^{(j)}$ für $j < i$ gilt. Diese Rückwirkung für Elemente $i \neq n$ sofort zu berechnen, ist nicht sinnvoll, denn ein späteres Element k mit $i < k < n$ kann wieder zu einer Änderung von allen Start- und Endgeschwindigkeiten führen. So bleibt die rückwärtige Ausbreitung der Geschwindigkeitsanpassungen zunächst unbeachtet und die Start- und Endgeschwindigkeiten werden für alle n Elemente berechnet. Im zweiten Berechnungsschritt wird die Bremsweglänge auf jedem Streckenelement betrachtet. Nun gilt es zu klären, ob der Weg $s_{ziel}^{(i)}$ ausreicht um von $v_{start}^{(i)}$ auf $v_{ziel}^{(i)}$ abzubremsen. Wenn $v_{ziel}^{(i)}$ auf dem Weg $s_{ziel}^{(i)}$ nicht erreicht werden kann, muss ermittelt werden, welchen Wert $v_{start}^{(i)}$ maximal annehmen darf. Der Bremsvorgang ist dabei gleichbedeutend mit einer Umkehr der Vorzeichen in den höheren zeitlichen Ableitungen der Geschwindigkeit. Somit können Start- und Endgeschwindigkeit des Elementes vertauscht werden, sodass $\tilde{v}_{start} = v_{ziel}$ und $\tilde{v}_{ziel} = v_{start}$ angenommen werden kann. Nun wird die „neue“ maximale Endgeschwindigkeit $\tilde{v}_{ziel_{neu}}^{(i)}$ mittels des Bahnplaners berechnet. Wenn $\tilde{v}_{ziel_{neu}}^{(i)}$ größer oder gleich v_{start} ist, genügt der vorhandene Weg s_{ziel} zum Bremsen. Kann auf dem betrachteten Element nicht mehr gebremst werden, ist nach dieser Berechnung zumindest die Startgeschwindigkeit $v_{start} = \tilde{v}_{ziel_{neu}}^{(i)}$ bekannt von welcher auf v_{ziel} abgebremst werden kann. Diese Berechnung müsste nun ausgehend vom aktuellen Element i für alle vorherigen Elemente $j < i$ durchgeführt werden, bis $\tilde{v}_{ziel_{neu}}^{(j)} \leq v_{start}^{(j)}$ eintritt. An dieser Stelle würde die Rückwirkung über die Elemente enden. Bei der Rückwärtsberechnung werden also alle Trajektorienelemente in umgekehrter Reihenfolge durchlaufen, wobei die Start- und Endgeschwindigkeiten gegeneinander getauscht werden. Auf diese „invertierte“ Trajektorie wird ebenfalls die Fallunterscheidung aus Abbildung 4.11 angewendet. Hierbei wird die Endgeschwindigkeit $\tilde{v}_{ziel}^{(i)}$ allerdings nur gesetzt, wenn $\tilde{v}_{ziel}^{(i)}$ kleiner $v_{start}^{(i)}$ ist.

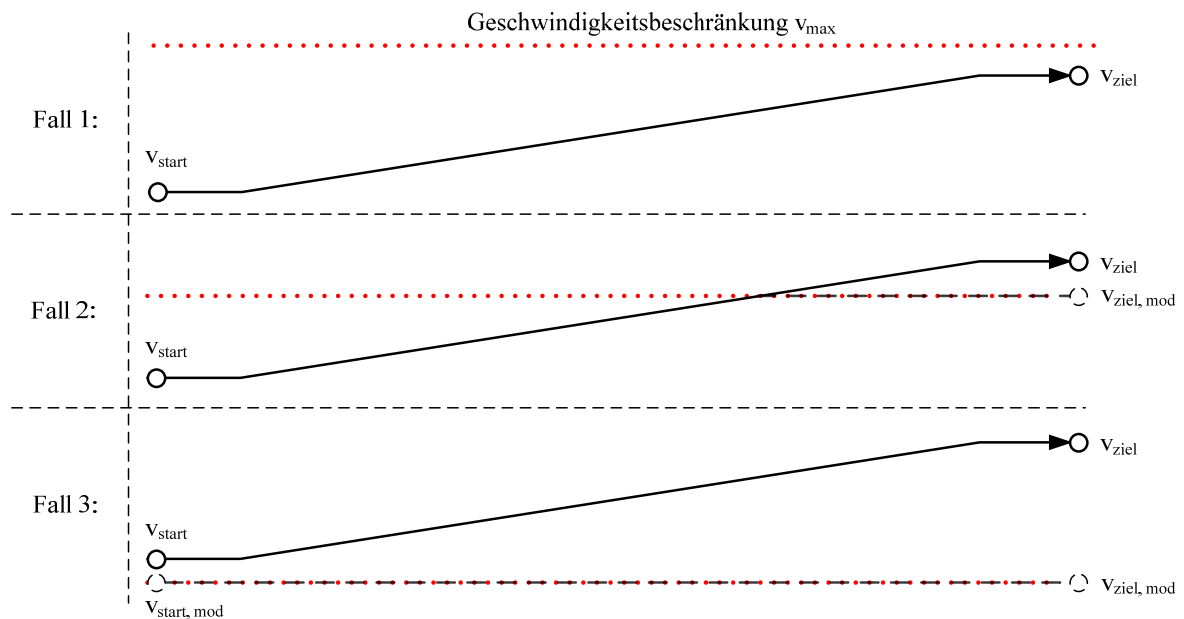


Abbildung 4.11: Mögliche Fälle bei der Bestimmung der Übergangsgeschwindigkeiten

Aus dem Minimum aller Übergangsgeschwindigkeiten der Vor- und Rückwärtsberechnung entsteht im dritten Schritt das gesuchte Geschwindigkeitsprofil (siehe Abbildung 4.12). Durch die vorgestellte Methode ist das Endergebnis nach maximal $2n$ Berechnungen bekannt. Somit wächst der Rechenaufwand, trotz der Rückwirkungen zwischen den Elementen, linear mit der Anzahl der Streckenelemente. Zusammenfassend kann festgestellt werden, dass mit dem vorgestellten Algorithmus eine kontinuierliche Bewegung über eine beliebige Anzahl von Basis-elementen geplant werden kann. Auch ist es möglich, Radialkräfte bei der Planung mit zu berücksichtigen, was im Besonderen bei der in dieser Arbeit betrachteten Anwendung entscheidend zu einer optimalen Regelgüte beiträgt.

In Abbildung 4.13 ist beispielhaft eine geplante Trajektorie dargestellt, welche aus mehreren Kreisbögen, Strecken sowie erweiterten Blosskurven besteht. In Anhang C sind die korrespondierenden kinematischen Profile dargestellt. Abbildung A 17 zeigt die berechnete Kinematik auf Bahnebene und Abbildung A 18 bis Abbildung A 20 die sich ergebenden dynamischen Sollwerte für die einzelnen Maschinenachsen.

Es ist deutlich zu erkennen, dass zu keinem Zeitpunkt die rot eingefärbten Beschränkungen verletzt werden. Zum besseren Verständnis des in diesem Abschnitt hergeleiteten Bahnplanungsalgorithmus ist in Abbildung A 21 der prinzipielle Ablauf dargestellt.

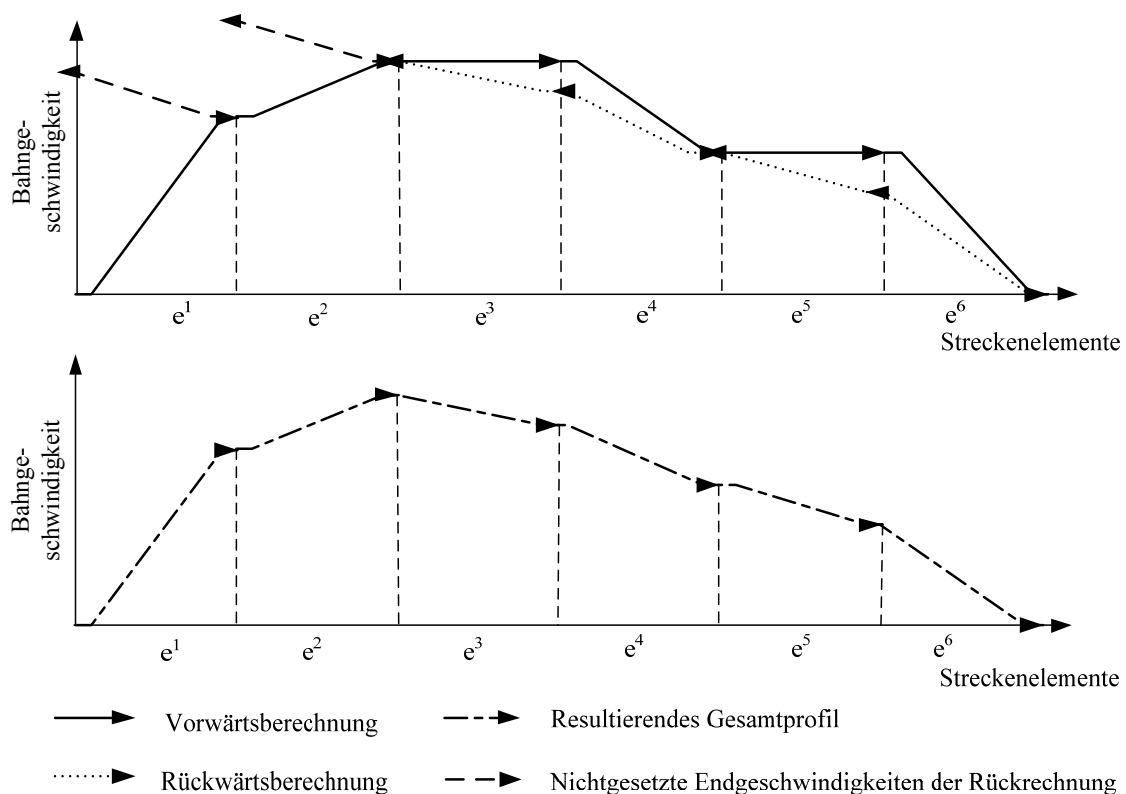


Abbildung 4.12: Schematische Darstellung zur Bestimmung der Elementübergangsgeschwindigkeiten

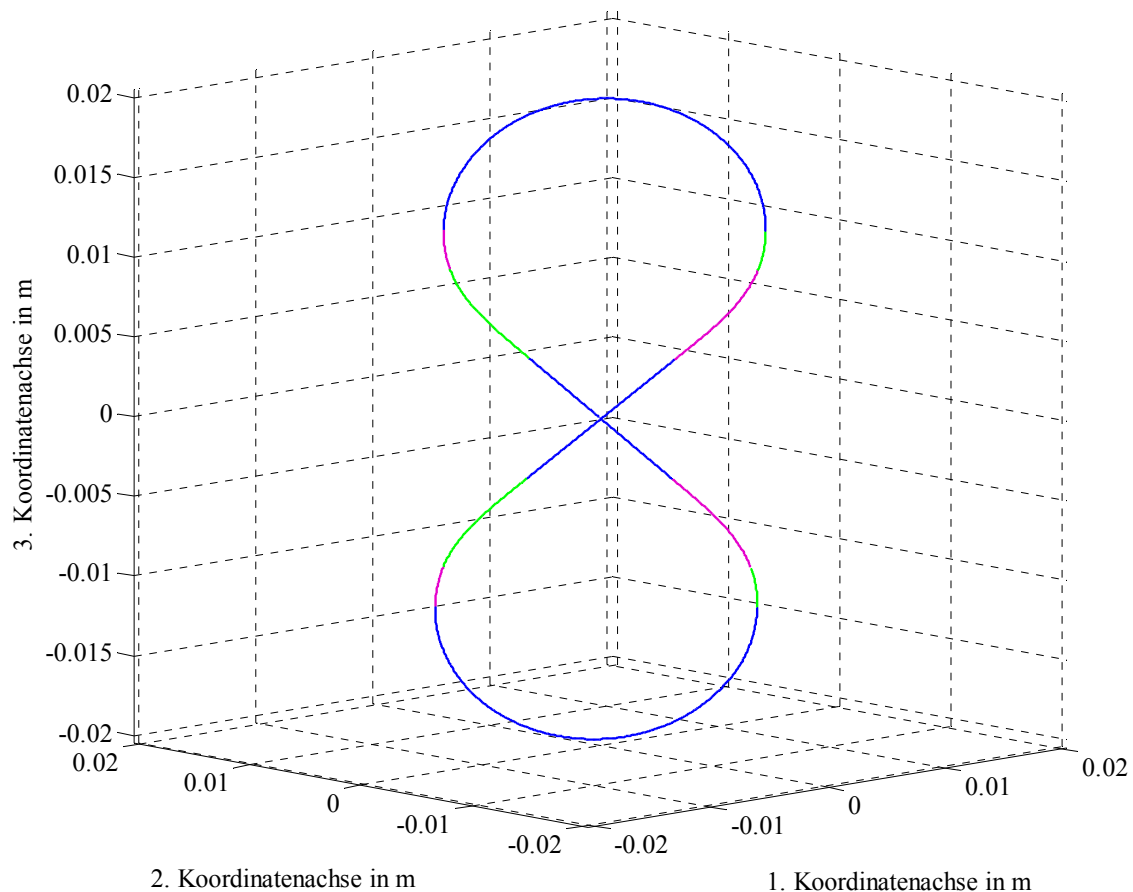


Abbildung 4.13: Verschlossene Trajektorie

4.2 Unterlagerte Regelung

Ein weiterer Bestandteil der vorgeschlagenen Trajektorienfolgeregelung ist die unterlagerte Regelung. Ihre Aufgabe ist die Ausregelung von exogenen Störungen sowie die Kompensation von systembedingten Modellierungsfehlern zwischen Vorsteuerung und dem zu regelnden System. Um den Einfluss der unterlagerten Regelung auf die Qualität der gesamten Trajektorienfolgeregelung untersuchen zu können, werden in den folgenden Abschnitten zwei unterschiedliche Regelungsansätze eingeführt.

Zum einen wird ein experimentell entworfener, zeitdiskreter PID-Regler eingesetzt und zum anderen kommt ein PI-Zustandsregler mit Führungsgrößenaufschaltung zur Anwendung. Dieser wird modellbasiert entworfen und damit fließt bedeutend mehr Systeminformation in die Reglerauslegung ein als in den Entwurf des PID-Reglers. Abgeschlossen wird der Abschnitt mit der Vorstellung eines speziellen Differentiationsalgorithmus, welcher es erlaubt, rauscharme Ableitungen mit einem minimalen Phasenverzug zu berechnen. Dieser Algorithmus ist für die unterlagerten Regler essentiell, da beide Reglertypen die nicht messbare Geschwindigkeit als Eingangsgröße verwenden.

4.2.1 PID-Regler

Als erste Variante kommt am xy-Demonstrator ein zeitdiskreter PID-Regler zum Einsatz. Dessen struktureller Aufbau ist in Abbildung 4.14 dargestellt. In diesem Blockschaltbild fällt auf, dass die Ableitung des Regelfehlers nicht durch eine einfache Differentiation sondern auf Basis der Soll- und Ist-Geschwindigkeit berechnet wird. Da die Ist-Geschwindigkeit messtechnisch in der geforderten Auflösung nicht erfasst werden kann, muss diese aus den Positionsmesswerten rekonstruiert werden. Neben einem Beobachterkonzept kommen hierzu spezielle Differentiationsalgorithmen in Frage und in der vorliegenden Arbeit wurde ein algebraischer Ableitungsschätzer eingesetzt (siehe Abschnitt 4.2.3). Dieser Aufwand ist notwendig, da die Messsignale der Interferometer von einem Rauschen von $\pm 0,5$ nm überlagert werden und somit eine einfache Differentiation zu unbrauchbaren Geschwindigkeitssignalen führt.

Aufgrund seiner einfachen Struktur wird der vorgestellte PID-Regler nach der Methode von Ziegler und Nichols bemessen [138].

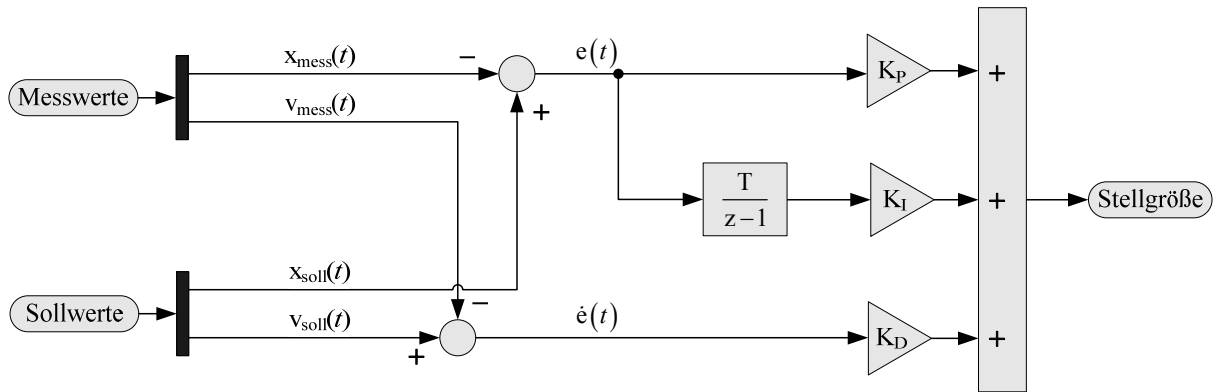


Abbildung 4.14: Struktur des zeitdiskreten PID-Reglers

4.2.2 PI-Zustandsregelung

Zur deutlichen Steigerung der Dynamik wird neben dem PID-Regler ein PI-Zustandsregler untersucht. In den folgenden beiden Abschnitten wird dieser Ansatz hergeleitet und darauf aufbauend ein modellbasierter Entwurf mittels Polvorgabe durchgeführt.

4.2.2.1 Entwurf mittels Polvorgabe

Für den Entwurf eines PI-Zustandsreglers nach [43, 88] wird das Konzept der klassischen Zustandsregelung um eine Ausgangsrückführung erweitert. Auf Basis der berechneten Regelabweichung kann der Zustandsregler um einen Proportional- sowie einen Integralanteil ergänzt werden. Eine schematische Darstellung des Ansatzes ist in Abbildung 4.15 zu sehen. Die Regelstrecke sowie der PI-Zustandsregler können durch die folgenden Differentialgleichungen beschrieben werden:

$$\dot{\mathbf{x}}(t) = \mathbf{A} \cdot \mathbf{x}(t) + \mathbf{B} \cdot \mathbf{u}(t) \quad (4.115)$$

$$\mathbf{y}(t) = \mathbf{C} \cdot \mathbf{x}(t) \quad (4.116)$$

$$\mathbf{u}(t) = -\mathbf{K} \cdot \mathbf{x}(t) + \mathbf{K}_p \cdot \dot{\mathbf{e}}(t) + \mathbf{K}_i \cdot \mathbf{e}(t) \quad (4.117)$$

$$\dot{e}(t) = w(t) - y(t) \quad (4.118)$$

Für $w = 0$ folgt nach dem Einsetzen von Gleichung (4.118) in Gleichung (4.117):

$$\dot{\mathbf{x}}(t) = \mathbf{A} \cdot \mathbf{x}(t) + \mathbf{B} \cdot \mathbf{u}(t) \quad (4.119)$$

$$\dot{e}(t) = -\mathbf{C} \cdot \mathbf{x}(t) \quad (4.120)$$

$$\mathbf{u}(t) = -\mathbf{K} \cdot \mathbf{x}(t) - \mathbf{K}_p \cdot \mathbf{C} \cdot \mathbf{x}(t) + \mathbf{K}_I \cdot e(t) \quad (4.121)$$

Nun kann das Zustandsraummodell des PI-Zustandsreglers abgeleitet werden:

$$\underbrace{\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{e}(t) \end{bmatrix}}_{\tilde{\mathbf{x}}(t)} = \underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{C} & \mathbf{0} \end{bmatrix}}_{\tilde{\mathbf{A}}} \cdot \underbrace{\begin{bmatrix} \mathbf{x}(t) \\ e(t) \end{bmatrix}}_{\tilde{\mathbf{x}}(t)} + \underbrace{\begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix}}_{\tilde{\mathbf{B}}} \cdot \underbrace{\mathbf{u}(t)}_{\tilde{\mathbf{u}}(t)} \quad (4.122)$$

$$\underbrace{\mathbf{u}(t)}_{\tilde{\mathbf{u}}(t)} = - \underbrace{\begin{bmatrix} \mathbf{K} + \mathbf{K}_p \cdot \mathbf{C} & -\mathbf{K}_I \end{bmatrix}}_{\tilde{\mathbf{K}}} \cdot \underbrace{\begin{bmatrix} \mathbf{x}(t) \\ e(t) \end{bmatrix}}_{\tilde{\mathbf{x}}(t)} \quad (4.123)$$

Die einzelnen Zustände und Matrizen dieses erweiterten Systems werden durch eine Tilde gekennzeichnet. Wird Gleichung (4.123) in Gleichung (4.122) eingesetzt, ergibt sich:

$$\dot{\tilde{\mathbf{x}}}(t) = \underbrace{(\tilde{\mathbf{A}} - \tilde{\mathbf{B}} \cdot \tilde{\mathbf{K}})}_{\tilde{\mathbf{A}}_R} \cdot \tilde{\mathbf{x}}(t) \quad (4.124)$$

$\tilde{\mathbf{A}}_R$ wird als Systemmatrix des geregelten Systems bezeichnet und die Eigenwerte beziehungsweise Polstellen dieser Matrix bestimmen das dynamische Verhalten des Gesamtsystems. Somit ergibt sich direkt die Proportionalverstärkung K_p aus der Zustandsraumdarstellung des zu regelnden Systems [43]:

$$\mathbf{K}_p = -\left(\tilde{\mathbf{C}} \cdot (\tilde{\mathbf{A}} - \tilde{\mathbf{B}} \cdot \tilde{\mathbf{K}})^{-1} \cdot \tilde{\mathbf{B}}\right)^{-1} = k_1 \quad (4.125)$$

Die Reglerparameter werden nach dem Verfahren der Polvorgabe bestimmt und die Berechnungsvorschriften sollen im Folgenden hergeleitet werden. Bei der Modellbildung für den Reglerentwurf wird davon ausgegangen, dass die modellbasierte Vorsteuerung (siehe Abbildung 4.1) nahezu alle Nichtlinearitäten aus dem betrachteten NPM-System eliminiert.

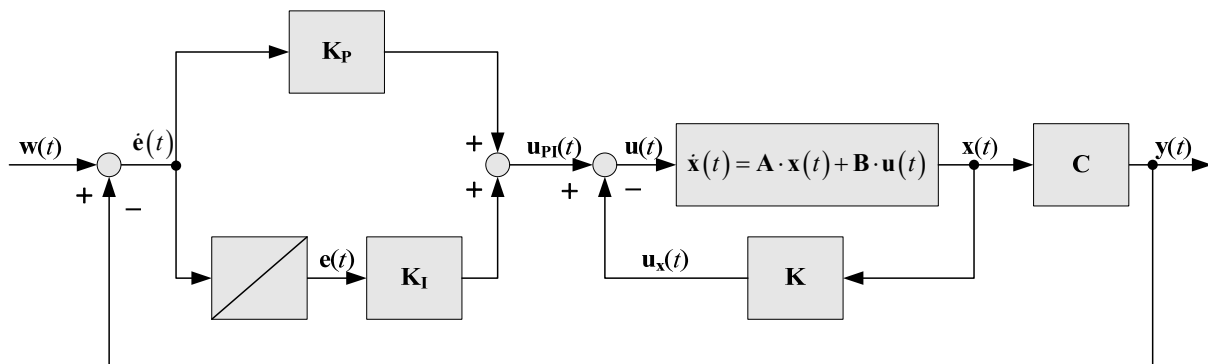


Abbildung 4.15: Struktur einer PI-Zustandsregelung

Demzufolge wird im Reglermodell lediglich der viskose Reibanteil $\sigma_2 \cdot \dot{x}(t)$ betrachtet. Analog zu Gleichung (3.12) kann das Verhalten einer Bewegungsachse beschrieben werden:

$$m \cdot \ddot{x}(t) + \sigma_2 \cdot \dot{x}(t) = F_A(t) \quad (4.126)$$

Im Zustandsraum gilt:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\dot{x}}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{\sigma_2}{m} \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_A \cdot k_V}{m} \end{bmatrix} \cdot u_s(t) \quad (4.127)$$

$$y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix}$$

Auf dieser Grundlage kann nun mithilfe von Gleichung (4.122) das Zustandsraummodell der zugehörigen PI-Zustandsregelung abgeleitet werden:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\dot{x}}(t) \\ \dot{e}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\frac{\sigma_2}{m} & 0 \\ -1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \\ e(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_A \cdot k_V}{m} \\ 0 \end{bmatrix} \cdot u_s(t) \quad (4.128)$$

$$y(t) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \\ e(t) \end{bmatrix}$$

Nach Gleichung (4.124) folgt für die Systemmatrix des geregelten Systems:

$$\tilde{\mathbf{A}}_R = (\tilde{\mathbf{A}} - \tilde{\mathbf{B}} \cdot \tilde{\mathbf{K}}) = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{k_A \cdot k_V \cdot (k_1 + K_P)}{m} & -\frac{\sigma_2}{m} - \frac{k_A \cdot k_V \cdot k_2}{m} & \frac{k_A \cdot k_V \cdot K_I}{m} \\ -1 & 0 & 0 \end{bmatrix} \quad (4.129)$$

Wie in Gleichung (4.129) zu erkennen ist, kann mithilfe der Reglerverstärkung $\tilde{\mathbf{K}} = [k_I \ k_2]$ direkt Einfluss auf die Eigenwerte der Systemmatrix genommen werden. Diese Eigenschaft macht sich das Verfahren der Polvorgabe zu Nutze und es wird gefordert:

$$\det(s \cdot \mathbf{I} - \tilde{\mathbf{A}}_R) \stackrel{!}{=} \prod_{i=0}^n (s - \lambda_i) \quad (4.130)$$

Die Parameter λ_i in Gleichung (4.130) stellen die gewünschten Eigenwerte der Systemmatrix $\tilde{\mathbf{A}}_R$ dar. Durch einen Koeffizientenvergleich können nun die Parameter in $\tilde{\mathbf{K}}$ bestimmt werden.

Für rein reelle Pole folgt aus Gleichung (4.130):

$$\det(s \cdot \mathbf{I} - \tilde{\mathbf{A}}_R) = (s - p_0) \cdot (s - p_1) \cdot (s - p_2) \quad (4.131)$$

Daraus ergeben sich die Reglerparameter K_I und \mathbf{K} :

$$K_{I_{reell}} = -\frac{p_0 \cdot p_1 \cdot p_2 \cdot m}{k_A \cdot k_V} \quad (4.132)$$

$$k_{1_{reell}} = \frac{m \cdot (p_0 \cdot p_1 + p_1 \cdot p_2 + p_0 \cdot p_2)}{2 \cdot k_A \cdot k_V} \quad (4.133)$$

$$k_{2_{reell}} = -\frac{m \cdot (p_0 + p_1 + p_2) + \sigma_2}{k_A \cdot k_V} \quad (4.134)$$

Im Falle eines konjugiert komplexen Polpaares ergibt sich aus Gleichung (4.130):

$$\det(s \cdot \mathbf{I} - \tilde{\mathbf{A}}_R) = (s - p_0) \cdot (s - (p_{re} + p_{im} \cdot i)) \cdot (s - (p_{re} - p_{im} \cdot i)) \quad (4.135)$$

und der Koeffizientenvergleich liefert für die entsprechenden Reglerverstärkungen:

$$K_{I_{imag}} = -\frac{p_0 \cdot (p_{re}^2 + p_{im}^2) \cdot m}{k_A \cdot k_V} \quad (4.136)$$

$$k_{1_{imag}} = \frac{m \cdot (2 \cdot p_0 \cdot p_{re} + p_{re}^2 + p_{im}^2)}{2 \cdot k_A \cdot k_V} \quad (4.137)$$

$$k_{2_{imag}} = -\frac{m \cdot (p_0 + 2 \cdot p_{re}) - \sigma_2}{k_A \cdot k_V} \quad (4.138)$$

4.2.2.2 Entwurf einer Vorsteuerung mit Führungsgrößenaufschaltung

Da der entworfene PI-Zustandsregler Teil einer Trajektorienfolgeregelung ist, wird zur weiteren Steigerung der Dynamik eine Vorsteuerung mit Führungsgrößenaufschaltung eingesetzt. Hierbei ist es das Ziel, dass im Idealfall die Ausgangsgröße und die Führungsgröße identisch sind. Weiterhin soll diese Beziehung auch für alle Ableitungen von $\mathbf{y}(t)$ beziehungsweise $\mathbf{w}(t)$ gelten. Wie Abbildung 4.16 zeigt, wird zur Berechnung des Vorfilters \mathbf{S} eine klassische Zustandsregelung ohne PI-Erweiterung angenommen. Die Ordnung der Strecke beträgt zwei und somit werden nur die ersten beiden Ableitungen der Führungsgröße mit berücksichtigt. Der Vektor \mathbf{S} besitzt daher die Form $\mathbf{S} = [K_{V0} \ K_{V1} \ K_{V2}]$. Demzufolge kann die Zustandsraumdarstellung des geregelten Systems wie folgt erweitert werden:

$$\dot{\mathbf{x}}(t) = (\mathbf{A} - \mathbf{B} \cdot \mathbf{K}) \cdot \mathbf{x}(t) + \mathbf{B} \cdot \mathbf{S} \cdot \begin{bmatrix} w(t) \\ \dot{w}(t) \\ \ddot{w}(t) \end{bmatrix} \quad (4.139)$$

$$\mathbf{y}(t) = \mathbf{C} \cdot \mathbf{x}(t) \quad (4.140)$$

Nach der Transformation des Systems in den Laplace-Bereich gilt für das geregelte System:

$$s \cdot \mathbf{X}(s) = (\mathbf{A} - \mathbf{B} \cdot \mathbf{K}) \cdot \mathbf{X}(s) + \mathbf{B} \cdot \mathbf{S} \cdot \begin{bmatrix} \mathbf{W}(s) \\ s \cdot \mathbf{W}(s) \\ s^2 \cdot \mathbf{W}(s) \end{bmatrix} \quad (4.141)$$

$$\mathbf{Y}(s) = \mathbf{C} \cdot \mathbf{X}(s) \quad (4.142)$$

Nun kann Gleichung (4.141) nach $\mathbf{X}(s)$ umgestellt werden:

$$\mathbf{X}(s) = (s \cdot \mathbf{I} - \mathbf{A} + \mathbf{B} \cdot \mathbf{K})^{-1} \cdot \mathbf{B} \cdot \mathbf{S} \cdot \begin{bmatrix} \mathbf{W}(s) \\ s \cdot \mathbf{W}(s) \\ s^2 \cdot \mathbf{W}(s) \end{bmatrix} \quad (4.143)$$

Anschließend wird (4.143) in (4.142) eingesetzt und es ergibt sich:

$$\mathbf{Y}(s) = \underbrace{\mathbf{C} \cdot (s \cdot \mathbf{I} - \mathbf{A} + \mathbf{B} \cdot \mathbf{K})^{-1} \cdot \mathbf{B} \cdot \mathbf{S}}_{G(s)} \cdot \underbrace{\begin{bmatrix} 1 \\ s \\ s^2 \end{bmatrix} \cdot \mathbf{W}(s)}_{U(s)} \quad (4.144)$$

Die Übertragungsfunktion des geregelten Systems $G(s)$ wird im Laplace-Bereich mit $b_0/N(s)$ beschrieben und so kann Gleichung (4.144) umgeformt werden zu:

$$\mathbf{Y}(s) = \frac{b_0}{N(s)} \cdot [K_{v0} \quad K_{v1} \quad K_{v2}] \cdot \begin{bmatrix} 1 \\ s \\ s^2 \end{bmatrix} \cdot \mathbf{W}(s) \quad (4.145)$$

Da gefordert wird, dass $\mathbf{Y}(s) = \mathbf{W}(s)$ gilt, kann Gleichung (4.145) dargestellt werden als:

$$1 = \frac{\mathbf{Y}(s)}{\mathbf{W}(s)} = \frac{b_0 \cdot K_{v0} + b_0 \cdot K_{v1} \cdot s + b_0 \cdot K_{v2} \cdot s^2}{a_0 + a_1 \cdot s + a_2 \cdot s^2} \quad (4.146)$$

Auch hier können die Parameter der Vorsteuerung mithilfe eines Koeffizientenvergleiches bestimmt werden. Unabhängig von der gewählten Pollage ergeben sich die Berechnungsvorschriften für den Vorfilter \mathbf{S} :

$$K_{v0} = k_1 \quad (4.147)$$

$$K_{v1} = \frac{\sigma_2}{k_A \cdot k_V} + k_2 \quad (4.148)$$

$$K_{v2} = \frac{m}{k_A \cdot k_V} \quad (4.149)$$

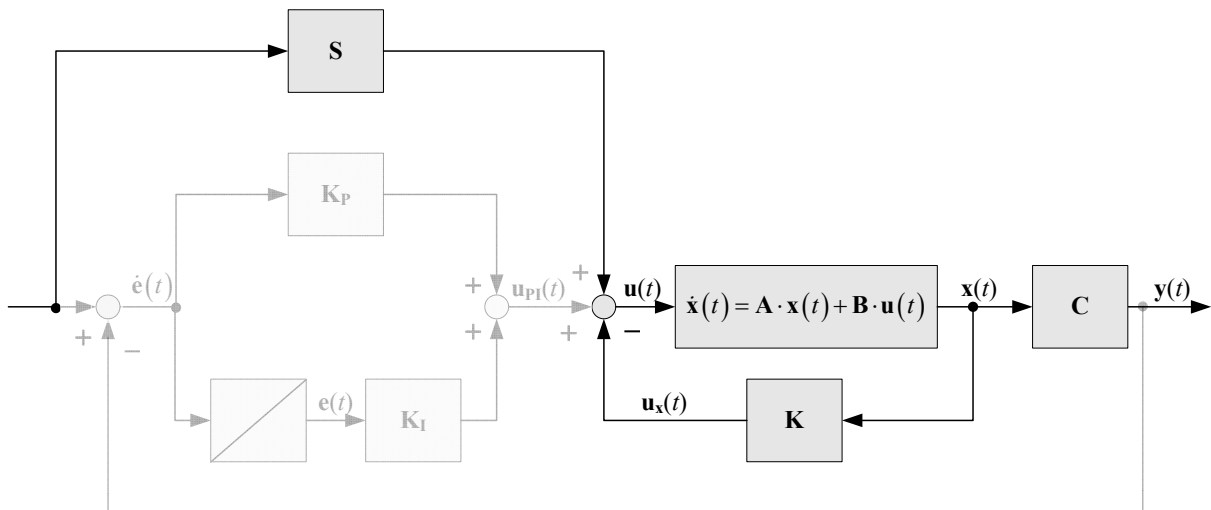


Abbildung 4.16: Erweiterung der PI-Zustandsregelung um einen Vorfilter mit Führungsgrößen-aufschaltung

4.2.3 Zustandsrekonstruktion

Beide vorgestellten Reglertypen benötigen die aktuelle Geschwindigkeit als Eingangsgröße. Da die Interferometer lediglich die Position hochgenau detektieren können und ein Geschwindigkeitsmesssystem mit der geforderten Auflösung nicht verfügbar ist, muss diese aus der gemessenen Position heraus rekonstruiert werden. Aufgrund der verrauschten Positionsmesswerte scheidet eine direkte Differentiation zur Lösung des Problems aus, denn der Rauschanteil im Signal wird durch ein solches Vorgehen unakzeptabel verstärkt. Mit Blick auf den Einsatzzweck eignen sich auch herkömmliche Filtertechniken nicht zur Qualitätsverbesserung des abgeleiteten Signals. Besonders die Zustandsregelung würde infolge des entstehenden Phasenverzuges instabil werden. In den Arbeiten von Reger *et al.* wird ein algebraischer Differentiationsalgorithmus (ADES) auf Basis von FIR-Filtern vorgestellt, welcher in Echtzeit nahezu phasenverzugsfreie und rauscharme Ableitungen eines beliebigen Eingangssignals berechnen kann [88, 99, 100, 136]. Da dieser *Ableitungsschätzer* Bestandteil des in diesem Kapitel vorgestellten modellbasierten Regelungskonzeptes ist, soll dessen Grundidee im Folgenden skizziert werden.

Es wird angenommen, dass das gemessene verrauschte Positionssignal $x(t)$ mit einer Taylorreihe der Ordnung N_T angenähert werden kann:

$$x(t) \approx x_{N_T}(t) = \sum_{o=0}^{N_T} \alpha_o \cdot \frac{t^o}{o!} \quad (4.150)$$

Die Koeffizienten α_o der Taylorreihe stellen dabei die Ableitungen des Positionssignals $x(t)$ zum Zeitpunkt $t = 0$ dar:

$$\alpha_o = \left. \frac{d^o x}{dt^o} \right|_{t=0} \triangleq x^{(o)} \quad (4.151)$$

Die Berechnung der Koeffizienten α_o gestaltet sich im Laplace-Bereich deutlich leichter als im Zeitbereich und daher wird eine Transformation in den Laplace-Bereich vorgenommen:

$$\mathcal{L} \left\{ x_{N_T}(t) = \sum_{o=0}^{N_T} \alpha_o \cdot \frac{t^o}{o!} \right\} = X_{N_T}(s) = \sum_{o=0}^{N_T} \frac{\alpha_o}{s^{o+1}} \quad (4.152)$$

Im nächsten Schritt wird Gleichung (4.152) derart umgeformt, dass ein beliebiger Koeffizient α_o aus dem Nutzsignal berechnet werden kann. Nach der Rücktransformation in den Zeitbereich folgt:

$$\alpha_o = \int_0^t \Pi_o(t, \tau) \cdot x_{N_T}(\tau) d\tau \quad (4.153)$$

Den Ausführungen in [137] folgend, kann die o -te Ableitung des Nutzsignals $x^{(o)}(t)$ an einem beliebigen Zeitpunkt t berechnet werden durch:

$$x^{(o)}(t) = (-1)^o \cdot \int_0^{T_{end}} \Pi_o(T_{end}, \tau) \cdot x(t - \tau) d\tau \quad (4.154)$$

Die Integrationsgrenze T_{end} stellt hierbei den Horizont des Nutzsignals dar, aus welchem die gewünschte Ableitung rekonstruiert wird. Bei der Wahl der Fensterbreite muss ein Kompromiss zwischen Rechenlast und Güte der Schätzung gefunden werden, denn je größer die Fensterbreite ist, desto kleiner wird die Eckfrequenz des Filters.

Die bisherige Herleitung betrachtet ein kontinuierliches Nutzsignal. Aufgrund der Tatsache, dass digitale Regelungssysteme nur zu äquidistanten Zeitpunkten ausgeführt werden [83], muss der vorgestellte Ansatz in eine zeitdiskrete Beschreibungsform überführt werden. Somit ergibt sich für die o -te Ableitung des Nutzsignals:

$$x^{(o)}(k) = \sum_{i=1}^{M_{ADES}} \Pi_{o,i} \cdot x_{k-i+1} \quad (4.155)$$

Wie in [137] gezeigt wird, sind die Gewichte $\Pi_{o,i}$ abhängig von der Ordnung der Taylorreihe N_T , der Fensterbreite $M_{ADES} = i \cdot T$ sowie einem Glättungsparameter ν . Alle diese Parameter ändern sich nicht während der Laufzeit, und so können die Gewichte $\Pi_{o,i}$ offline bestimmt werden. Lediglich Gleichung (4.155) muss in jedem Tastschritt ausgeführt werden.

Die Berechnungsvorschrift zur Bestimmung der Gewichte lautet [137]:

$$\Pi_{o,i} = \frac{(N_T + o + \nu + 1)! (N_T + 1)!}{(M_{ADES} T)^o} \sum_{\kappa_1=0}^{N_T-o} \sum_{\kappa_2=0}^o \sum_{\kappa_3=0}^{\nu+\kappa_1+\kappa_2} \dots \quad (4.156)$$

$$\left[\frac{(-1)^{N_T+\kappa_1-\kappa_2-\kappa_3} \left(\left(\frac{k}{M_{ADES}} \right)^{N_T+\kappa_1-\kappa_2-\kappa_3+1} - \left(\frac{k-1}{M_{ADES}} \right)^{N_T+\kappa_1-\kappa_2-\kappa_3+1} \right)}{\kappa_1! \kappa_2! \kappa_3! (N_T - o - \kappa_1)! (o - \kappa_2)! (N_T - \kappa_1 - \kappa_2)! (\nu - \kappa_3 + \kappa_1 + \kappa_2)!} \dots \right]$$

$$\left[\frac{1}{(N_T - \kappa_1 + 1)(N_T + \kappa_3 - \kappa_1 - \kappa_2 + 1)} \right]$$

Um die Leistungsfähigkeit des vorgestellten algebraischen Ableitungsschätzers zu zeigen, ist in Abbildung 4.17 die erste Ableitung eines verrauschten periodischen Signals zu sehen. Das sinusförmige Nutzsignal mit einer Amplitude von 30 nm wurde mit weißem Rauschen ($\sigma^2 = 5 \text{ nm}^2$) überlagert und anschließend sowohl mit DT₁-Glieder ($f_g = 10 \text{ Hz}$) als auch mit dem vorgestellten algebraischen Ableitungsschätzer (ADES) differenziert.

Es ist klar zu erkennen, dass ADES dem gewöhnlichen DT₁-Glieder bezüglich Rauschunterdrückung und Phasenverzug überlegen ist. Mit dem Einsatz des algebraischen Ableitungsschätzers wird es daher möglich, ein qualitativ hochwertiges Geschwindigkeitssignal aus den gemessenen Positionswerten zu rekonstruieren und dieses in der unterlagerten Regelung mit einzusetzen.

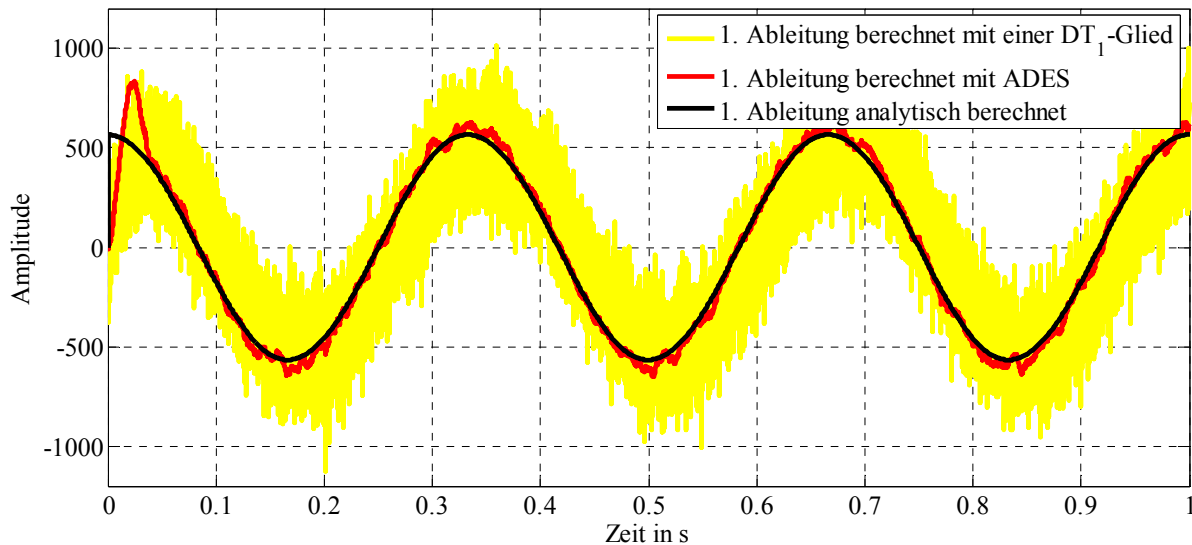


Abbildung 4.17: Leistungsfähigkeit des eingesetzten algebraischen Ableitungsschätzers

4.3 Adaptive Vorsteuerung

Kernkomponente der Trajektorienfolgeregelung aus Abbildung 4.1 ist die Vorsteuerung. Im Abschnitt 3.4 konnte gezeigt werden, dass eine dynamische Modellierung der Reibkraft das Reibverhalten in den Kugelführungen des xy-Demonstrators sehr gut nachbilden kann. Somit bietet es sich aus regelungstechnischer Sicht an, den dominanten Einfluss (siehe Abschnitt 3.4) der Reibkraft modellbasiert vorherzusagen und vorsteuernd zu kompensieren. Erste experimentelle Versuche der Reibkraftkompensation mit einem statischen Reibmodell scheiterten, denn wie schon in Abschnitt 3.4.3.3 gezeigt wurde, variiert die Reibkraft bei dieser Art von NPM-Systemen sehr stark mit der Zeit sowie dem Ort. Um dieses Problem zu lösen, wird eine Nachführung der Modellparameter in Echtzeit vorgeschlagen, denn nur so kann die nötige Flexibilität und Anpassungsfähigkeit des Regelungssystems erreicht werden [9, 10]. Die aus dieser Vorgehensweise resultierende Systemstruktur ist beispielhaft für eine Bewegungsachse des xy-Demonstrators in Abbildung 4.18 dargestellt. Die Parameternachführung kann erreicht werden, indem das verwendete Reibmodell zweimal in das dynamische Regelungssystem integriert wird. Zum einen fungiert das inverse Systemmodell als Vorsteuerung und zum anderen wird das identische Modell parallel zum realen System betrieben, um eine Bewertung der Modellgüte zu realisieren. In der konkreten Anwendung wird das eingesetzte Reibmodell zweimal berechnet.

Zu Beginn wird $u_V(k)$ mithilfe eines gültigen Modellparametervektors $\Psi(k)$ geschätzt und dadurch die Reibkraft vorsteuernd kompensiert. Anschließend wird die Reaktion des Systems messtechnisch erfasst und auf Basis dieser Information der Modellfehler $e_M(k)$ bestimmt.

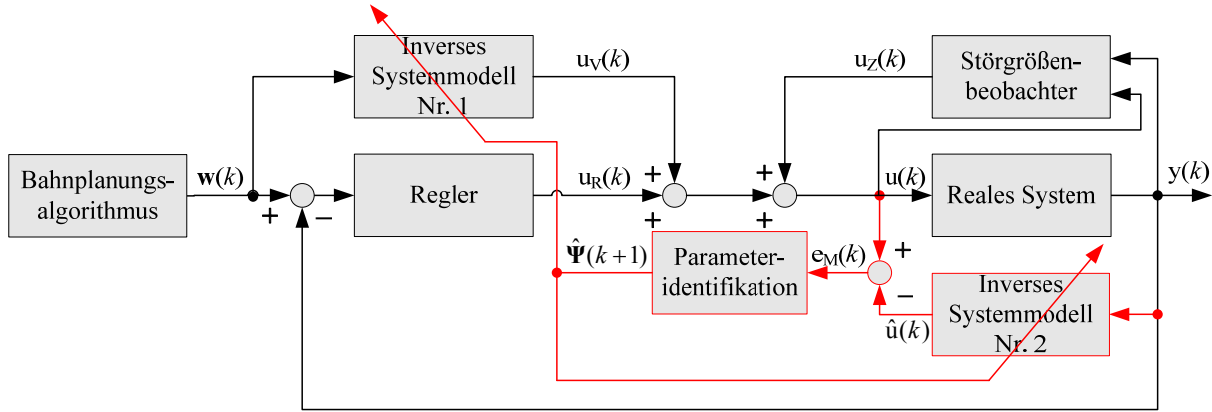


Abbildung 4.18: Erweiterte Regelungsstruktur mit adaptiver Vorsteuerung

Der für den nächsten Tastschritt gültige Modellparametervektor $\hat{\Psi}(k+1)$ wird nun durch einen Schätzalgorithmus derart angepasst, sodass der Modellfehler minimal wird. Der beschriebene Vorgang wird in jedem Tastschritt wiederholt.

Im Abschnitt 4.3.1 wird das so erweiterte adaptive DNLRX-Modell (ADNLRX-Modell) vorgestellt. Hierbei wird im Besonderen auf die Funktionsweise des genutzten Parameterschätzverfahrens eingegangen. Die Struktur des adaptiven vollvermaschten neuronalen Netzes wird in Abschnitt 4.3.2 näher erläutert. Auch hier wird der Schwerpunkt des Abschnittes auf der Darstellung des verwendeten echtzeitfähigen Lernalgorithmus liegen.

4.3.1 Adaptives DNLRX Modell

Um die Echtzeitfähigkeit der Parameterschätzung zu gewährleisten werden nur die linearen Parameter des DNLRX-Modells nachgeführt. Der nichtlineare Schwellwertvektor Δ wird, wie in Abschnitt 3.4.3.3 beschrieben, während des Parameteridentifikationsprozesses bestimmt und bleibt während der gesamten Laufzeit unverändert.

Weiterhin stellte sich in ausgiebigen experimentellen Untersuchungen heraus, dass eine Adaption des linearen Parametervektors Θ_i in Gleichung (3.39) die Güte der Reibkraftvorhersage nur geringfügig beeinflusst. In Folge dessen werden die Parameter in Θ_i auch nicht adaptiert, sondern es wird die gesamte gewichtete Summe $\sum_{i=0}^{n_\delta} \Theta_i^T \cdot \delta(k-i)$ aus Gleichung (3.39)

mit einem adaptiven Parameter $\varsigma(k)$ multipliziert. Somit ergibt sich eine neue Berechnungsvorschrift für das ADNLRX-Modell:

$$F_A(k) = \sum_{i=0}^{n_x} g_i(k) \cdot x(k-i) + \varsigma(k) \cdot \sum_{i=0}^{n_\delta} \Theta_i^T \cdot \delta(k-i) + b(k) \quad (4.157)$$

Abgesehen von der Gewichtung des nichtlinearen Modellanteiles mit ς werden in jedem Tastschritt die linearen Parameter g_i neu geschätzt. Somit kann der zu adaptierende Parametervektor $\Psi(k)$ wie folgt definiert werden:

$$\Psi(k) \triangleq [g_0(k) \dots g_n(k) : \varsigma(k) : b(k)]^T \quad (4.158)$$

Weiterhin muss, um den systemspezifischen Anforderungen von NPM-Systemen Rechnung zu tragen, neben der Nachführung der linearen Modellparameter das herkömmliche ADNLRX-Modell weiter angepasst werden. Aufgrund der hohen Abtastrate liegen die Positionsmesswerte zeitlich sehr nah beieinander. Mit dem Ziel, einen größeren Zeithorizont abdecken zu können, wird daher eine ganzzahlige Schrittweite α_{ADNLRX} zwischen den vom ADNLRX-Modell verwendeten Messwerten eingeführt. Gleichung (4.157) verändert sich damit zu:

$$F_A(k) = \sum_{i=0}^{n_x} g_i(k) \cdot x(k - \alpha_{ADNLRX} \cdot i) + \zeta(k) \cdot \sum_{i=0}^{n_\delta} \Theta_i^T \cdot \delta(k - i) + b(k) \quad (4.159)$$

Echtzeitfähiges Parameterschätzverfahren

In Anbetracht der Tatsache, dass pro Tastschritt dem Regelungsalgorithmus lediglich 100 μ s Rechenzeit zur Verfügung stehen, können nur Verfahren mit einer sehr geringen Rechenkomplexität eingesetzt werden. Das einzige Verfahren, welches den harten Echtzeitanforderungen genügt, ist eine rekursive Variante der Methode der kleinsten Quadrate (MKQ), auch rekursive Regression genannt [132]. Mit diesem Verfahren können natürlich nur Parameter bestimmt werden, welche linear in das Modell eingehen. Dieses Verfahren soll im Folgenden kurz vorgestellt werden. Bei einem allgemeinen parameterlinearen Modell der Form

$$\begin{aligned} \hat{y}(k) &= \mathbf{m}(k)^T \cdot \hat{\Psi} \quad \text{mit} \\ \mathbf{m}(k) &= [m_1(k) \dots m_r(k)]^T \\ \hat{\Psi} &= [\hat{\psi}_1 \dots \hat{\psi}_r]^T \end{aligned} \quad (4.160)$$

kann die Identifikationsaufgabe mit der Methode der kleinsten Quadrate gelöst werden. Zur Parameterbestimmung bedarf es der Beobachtung des Systemeingangsvektors $\mathbf{m}(k)$ sowie des Systemausgangs $y(k)$. Mithilfe dieser Daten kann der Vektor des Modellfehlers $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ gebildet und auf dessen Basis eine quadratische Gütefunktion formuliert werden:

$$\begin{aligned} J = \mathbf{e}^T \cdot \mathbf{e} &= [\mathbf{y} - \mathbf{M} \cdot \hat{\Psi}]^T \cdot [\mathbf{y} - \mathbf{M} \cdot \hat{\Psi}] \quad \text{mit} \\ \mathbf{M} &= [\mathbf{m}(0) \dots \mathbf{m}(N-1)]^T \\ \mathbf{y} &= [\mathbf{y}(0) \dots \mathbf{y}(N-1)]^T \end{aligned} \quad (4.161)$$

Dabei bezeichnet N die Anzahl der Beobachtungen. Bei r Parametern muss für eine korrekte Schätzung $N \geq r$ gelten, da sonst das entstehende Gleichungssystem unterbestimmt wäre. Die Modellparameter $\hat{\Psi}$ können durch Ableitung der Gütefunktion nach den Modellparametern und anschließendem Nullsetzen analytisch berechnet werden:

$$\frac{\partial J}{\partial \hat{\Psi}} = -2 \cdot \mathbf{M}^T \cdot \mathbf{y} + 2 \cdot \mathbf{M}^T \cdot \mathbf{M} \cdot \hat{\Psi} = \mathbf{0} \quad (4.162)$$

Wird Gleichung (4.162) nach dem zu bestimmenden Parametervektor umgestellt, folgt:

$$\hat{\Psi} = \mathbf{P} \cdot \mathbf{M}^T \cdot \mathbf{y} \quad \text{mit} \quad (4.163)$$

$$\mathbf{P} = [\mathbf{M}^T \cdot \mathbf{M}]^{-1}$$

Hierbei kann \mathbf{P} als die Kovarianzmatrix des Schätzfehlers interpretiert werden [61]. Gleichung (4.163) repräsentiert somit die analytische Lösung des Optimierungsproblems und wird als Regressionsgleichung bezeichnet. Die rekursive Variante der Methode der kleinsten Quadrate, in der Fachliteratur kurz als „Recursive least squares“ (RLS) bezeichnet, beruht auf dem bereits beschriebenen Verfahren der MKQ. Dabei wird angenommen, dass zum Zeitpunkt der Schätzung nicht alle Beobachtungen vorliegen, sondern mit jedem Zeitschritt ein neuer Messwert hinzukommt. Auf Grundlage dieser neuen Information wird nun zu jedem Zeitschritt der Parametervektor $\hat{\Psi}$ weiter verbessert. Ausgehend von Gleichung (4.163) berechnet sich der geschätzte Parametervektor zum Zeitpunkt k aus:

$$\hat{\Psi}(k) = \mathbf{P}(k) \cdot \mathbf{M}^T(k) \cdot \mathbf{y}(k) \quad \text{mit}$$

$$\mathbf{P}(k) = [\mathbf{M}^T(k) \cdot \mathbf{M}(k)]^{-1} \quad (4.164)$$

$$\mathbf{M}(k) = [\mathbf{m}(1) \dots \mathbf{m}(k)]^T$$

$$\mathbf{y}(k) = [\mathbf{y}(1) \dots \mathbf{y}(k)]^T$$

Zum Zeitpunkt $k+1$ stehen zusätzlich die Werte des Systemeingangs $\mathbf{m}(k+1)$ und des Systemausgangs $\mathbf{y}(k+1)$ zu Verfügung. Somit gilt nun für den Parametervektor im nächsten Zeitschritt:

$$\hat{\Psi}(k+1) = \mathbf{P}(k+1) \cdot \mathbf{M}^T(k+1) \cdot \mathbf{y}(k+1) \quad (4.165)$$

Durch Einbeziehung der Schätzgleichung zum Zeitpunkt k ergibt sich:

$$\hat{\Psi}(k+1) = \mathbf{P}(k+1) \cdot \begin{bmatrix} \mathbf{M}(k) \\ \mathbf{m}^T(k+1) \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{y}(k) \\ \mathbf{y}(k+1) \end{bmatrix} \quad (4.166)$$

$$= \mathbf{P}(k+1) \cdot [\mathbf{M}^T(k) \cdot \mathbf{y}(k) + \mathbf{m}^T(k+1) \cdot \mathbf{y}(k+1)]$$

Mit Gleichung (4.166) steht nun eine rekursive Beziehung zur Berechnung des Parametervektors zur Verfügung. Dieser Ausdruck kann noch weiter vereinfacht werden, und es ergeben sich die endgültigen Gleichungen für die rekursive Regression [61, 132]:

$$\mathbf{K}_{\text{RLS}}(k+1) = \mathbf{P}(k) \cdot \mathbf{m}(k+1) \cdot [1 + \mathbf{m}^T(k+1) \cdot \mathbf{P}(k) \cdot \mathbf{m}(k+1)]^{-1} \quad (4.167)$$

$$\hat{\Psi}(k+1) = \hat{\Psi}(k) + \mathbf{K}_{\text{RLS}}(k+1) \cdot [y(k+1) - \mathbf{m}^T(k+1) \cdot \hat{\Psi}(k)] \quad (4.168)$$

$$\mathbf{P}(k+1) = \mathbf{P}(k) - \mathbf{K}_{\text{RLS}}(k+1) \cdot \mathbf{m}^T(k+1) \cdot \mathbf{P}(k) \quad (4.169)$$

Für eine rekursive Berechnung des Parametervektors müssen Anfangswerte für die zu bestimmenden Parameter $\hat{\Psi}(k=0)$ sowie die Fehlerkovarianzmatrix $\mathbf{P}(0)$ bekannt sein.

Häufig fehlen diese Startwerte, sodass eine Verwendung allgemein gültiger Startwerte vorgenommen werden muss. In [61] wird gezeigt, dass bei einer Initialisierung $\mathbf{P}(0) = c \cdot \mathbf{I}$ mit großen Werten für c die Werte von $\mathbf{P}(0)$ einen vernachlässigbar kleinen Einfluss auf das rekursiv berechnete $\mathbf{P}(k)$ haben. In der praktischen Anwendung haben sich $100 \leq c \leq 1000$ und $\mathbf{P}(0) = \mathbf{0}$ bewährt. Bei näherer Betrachtung der Gleichungen (4.167) bis (4.169) fällt auf, dass eine permanente Nachführung der Modellparameter nicht möglich ist, denn die Elemente der Matrix $\mathbf{P}(k)$ streben mit wachsender Datenanzahl gegen Null. Dies führt nach Gleichung (4.167) zu einer Korrekturmatrix $\mathbf{K}_{\text{RLS}} = \mathbf{0}$ und damit stagniert die Nachführung des Parametervektors nach einer bestimmten Zeitspanne. Zur Lösung dieses Problems ist es notwendig, ein Konvergieren der Elemente der Matrix $\mathbf{P}(k)$ gegen Null zu verhindern. Dies kann entweder durch Addition einer Korrekturmatrix \mathbf{P}^* oder durch Division mit einer konstanten Vergessensrate ξ ($0 < \xi < 1$) erreicht werden. Am gebräuchlichsten ist die Verwendung der Vergessensrate ξ , welche zu einer exponentiellen Gewichtung der für die Schätzung benutzten Daten führt. Somit muss Gleichung (4.169) erweitert werden zu:

$$\mathbf{P}(k+1) = [\mathbf{P}(k) - \mathbf{K}_{\text{RLS}}(k+1) \cdot \mathbf{m}^T(k+1) \cdot \mathbf{P}(k)] / \xi \quad (4.170)$$

Nach der erfolgreichen Implementierung wurde der konventionelle RLS-Schätzalgorithmus anhand eines Messwertdatensatzes validiert. Dafür wurde ein Reibmodell (ADNLRX-Modell der Ordnung $M = 7$, $n_x = 7$, $n_\delta = 0$) mithilfe aufgenommener Messwerte identifiziert und simulativ verifiziert. Wie Abbildung 4.19 zeigt, neigt der RLS-Algorithmus zu Instabilitäten. Es ist deutlich zu erkennen, dass sich der konventionelle RLS-Algorithmus nach circa 12 s destabilisiert. Grund für dieses Verhalten sind die relativ stark verrauschten Messgrößen, die dem Schätzalgorithmus als Datenbasis dienen (siehe Abbildung 4.18). Zur Lösung des Problems wird in der vorliegenden Arbeit eine numerisch stabilere Variante des RLS-Algorithmus eingesetzt. Die Wahl fällt hierbei auf die so genannte U-D-Faktorisierung der Fehlerkovarianzmatrix $\mathbf{P}(k)$, welche 1975 von Gerald J. Bierman vorgestellt wurde [23]. Die Eigenschaften dieses Algorithmus werden in [22] detailliert untersucht und im Folgenden soll lediglich die Idee etwas näher beleuchtet werden.

Bei der U-D-Faktorisierung wird die Matrix $\mathbf{P}(k)$ in eine obere Dreiecksmatrix $\mathbf{U}(k)$ und eine Diagonalmatrix $\mathbf{D}(k)$ zerlegt:

$$\mathbf{P}(k) = \mathbf{U}(k) \mathbf{D}(k) \mathbf{U}^T(k) \quad (4.171)$$

Vorteilhaft an dieser Faktorisierung ist die erhöhte numerische Stabilität und Konvergenz, denn Berechnungen mit Dreiecksmatrizen erfordern weniger arithmetische Operationen. Ausgehend von Gleichung (4.171) entwickelte Bierman Rekursionsgleichungen für $\mathbf{U}(k+1)$ und $\mathbf{D}(k+1)$, die so in den RLS-Algorithmus eingebettet werden, dass eine explizite Berechnung von $\mathbf{P}(k+1)$ entfällt.

Der Validierungsprozess aus Abbildung 4.19 wurde auch unter Nutzung des Algorithmus von Bierman durchgeführt.

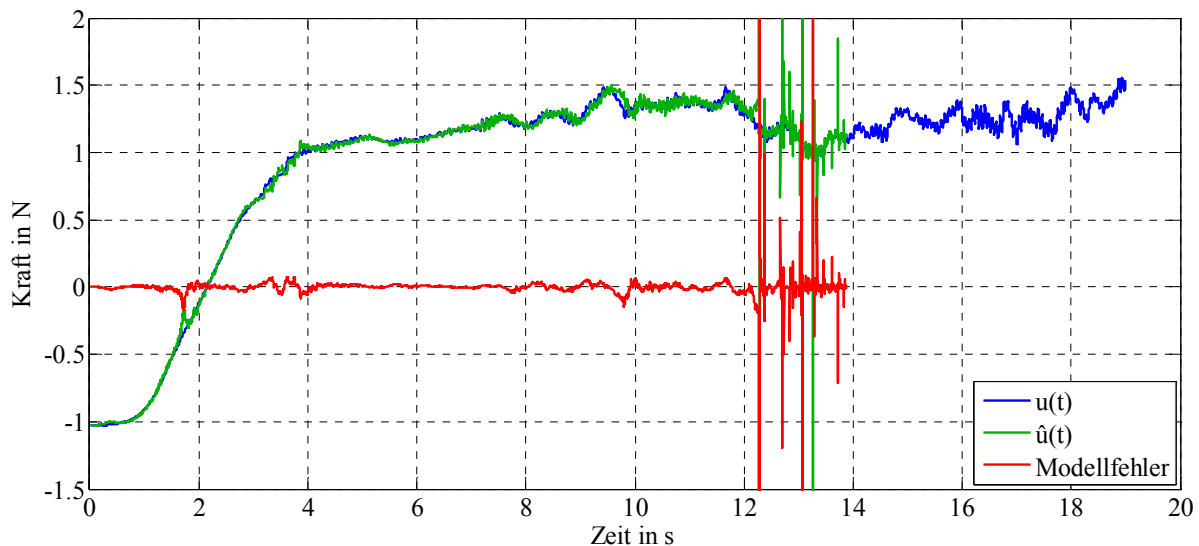


Abbildung 4.19: Rekursive Parameterschätzung mittels konventionellem RLS-Schätzalgorithmus

Es ist in Abbildung 4.20 klar zu erkennen, dass der Algorithmus mit U-D-Zerlegung über den gesamten Simulationshorizont stabil bleibt und keine Instabilitäten bei der Parameterschätzung auftreten.

Neben den Stabilitätsuntersuchungen wurde im Zuge der Leistungsbewertung auch die Laufzeit (@ 2,0 GHz Intel Core2Duo) beider Schätzverfahren analysiert. Es kamen Datensätze mit jeweils 2000 Messpunkten zum Einsatz. Die in Abbildung 4.21 dargestellten Zeiten entsprechen den Mittelwerten aus jeweils 200 Identifikationsdurchläufen.

Besonders bei einer geringen Anzahl zu schätzender Parameter ist die U-D-Faktorisierung neben der erhöhten numerischen Stabilität auch im Bezug auf die Rechenkomplexität dem konventionellen RLS-Algorithmus überlegen.

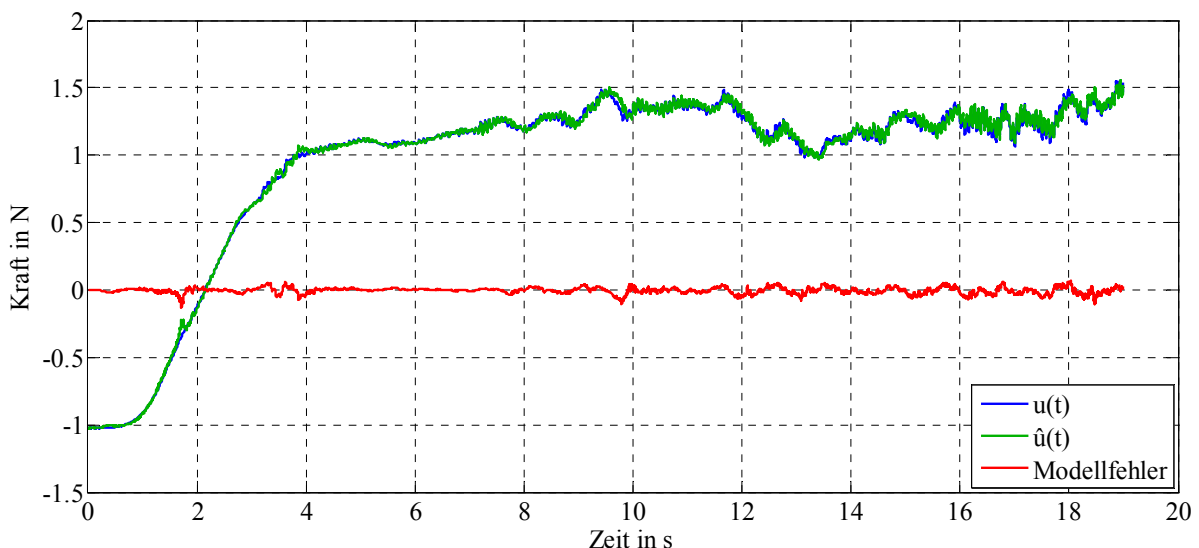


Abbildung 4.20: Rekursive Parameterschätzung mittels RLS-Schätzalgorithmus mit U-D-Faktorisierung

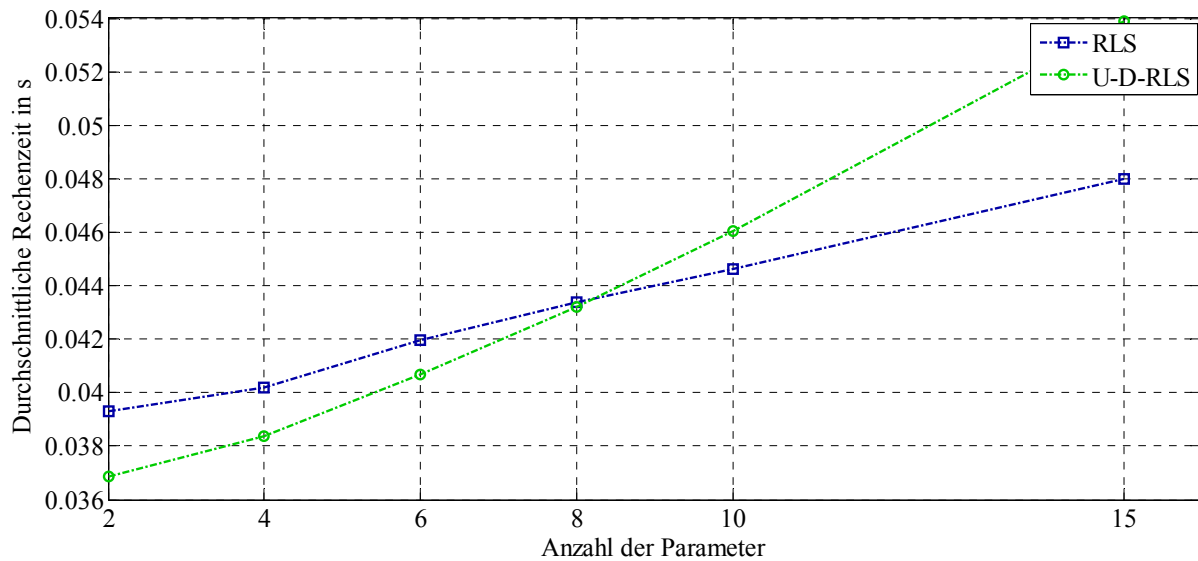


Abbildung 4.21: Gegenüberstellung der Laufzeiten beider RLS-Algorithmen

4.3.2 Adaptives neuronales Netz

Als Alternative zum DNLRX-Modell wird das vollvermaschte künstliche neuronale Netz zur Kompensation des nichtlinearen Systemverhaltens eingesetzt. Auch das neuronale Netz muss adaptiv gestaltet werden, um das stark schwankende Systemverhalten nachbilden zu können. Hierzu müssen alle Gewichte des Netzes während der Laufzeit adaptiert werden und dies erfordert einen echtzeitfähigen Lernalgorithmus.

Echtzeitfähiges Lernverfahren

Zu Beginn der Arbeiten wurde der schon bei der Identifikation erfolgreich eingesetzte Levenberg-Marquardt-Algorithmus favorisiert. Während der Erprobung zeigte sich jedoch, dass der Algorithmus aufgrund seiner Rechenkomplexität nicht in 100 μ s ausgeführt werden kann. Als Alternative wird daher der klassische „Backpropagation“ Algorithmus (BPA) eingesetzt, denn nur dieser erwies sich auf der eingesetzten Rechenhardware als echtzeitfähig. Die prinzipielle Funktionsweise des BPA wird nun im Folgenden kurz skizziert.

Prinzipiell lässt sich der BPA in drei Berechnungsschritte unterteilen. Im ersten Schritt wird die Netzausgabe für einen Eingangsdatensatz berechnet. Daraufhin wird im zweiten Schritt auf Basis des Ausgangsfehlers die Änderung aller Gewichte im KNN bestimmt. Im letzten Schritt wird der Lernerfolg ermittelt und in dessen Abhängigkeit die Schrittweite für die nächste Gewichts Anpassung berechnet. Der Algorithmus endet, wenn ein bestimmtes Zielkriterium erfüllt ist.

Grundlage des BPA ist analog zum Levenberg-Marquardt-Algorithmus der Gradientenabstieg in einem Gütegebirge. Die Dimension dieses Gebirges ist abhängig von der Anzahl der Netzeingänge, der Netzausgänge und der Gewichte im KNN. Als Fehlerfunktion wird auch hier der quadratische Fehler zwischen Netzausgang und gewünschtem Ausgang verwendet. Die Fehlerfunktion für den k -ten Musterdatensatz kann wie folgt formuliert werden:

$$E(k) = (y(k) - y_{net}(k))^2 \quad (4.172)$$

Da je Lernepoche eine bestimmte Menge n an Musterdatensätzen betrachtet wird, gilt:

$$E = \frac{1}{n} \sum_{k=1}^n E(k) \quad (4.173)$$

Der Ausgangsfehler E wird, wie bereits in Abschnitt 3.4.3.4 ausführlich dargelegt, durch das neuronale Netz propagiert und der Gewichtsgradient in der Ausgangsschicht wird ermittelt durch:

$$\begin{aligned} w_{\Delta}^O &= \frac{\partial E(k)}{\partial w_i^O} \\ &= -2 \cdot (y - y_{net}) \cdot f^H(net_i^H) \end{aligned} \quad (4.174)$$

Ebenso können die Gewichtsgradienten der Hiddenschicht berechnet werden aus:

$$\begin{aligned} w_{\Delta}^H &= \frac{\partial E(k)}{\partial w_{i,q}^H} \\ &= -2 \cdot (y - y_{net}) \cdot \sum_{i=1}^N w_i^O \cdot f^{H'}(net_i^H) \cdot U(k) \end{aligned} \quad (4.175)$$

Mit Gleichung (4.174) und (4.175) stehen nun Berechnungsvorschriften für die Gradienten aller Gewichte im KNN zur Verfügung. Basierend auf den Gradienten kann die Vorschrift zur Anpassung eines beliebigen Gewichtes w formuliert werden:

$$w^{neu} = w^{alt} - (mc \cdot \Delta w^{alt} - (1 - mc) \cdot l_r \cdot w_{\Delta}) \quad (4.176)$$

Hierbei ist w_{Δ} der berechnete Gradient des betrachteten Neurons und mc ein so genannter Momentumterm. Durch mc ($0 < mc < 1$) wird eine Begrenzung der erlaubten Richtungsänderung des Gradienten bewirkt. Somit wird ein Oszillieren des Gradienten in schmalen Schluchten des Gütegebirges verhindert und die Konvergenz des Verfahrens erhöht.

Wie auch schon beim Training des Netzes wird nach der Anpassung aller Gewichte überprüft, ob der Ausgangsfehler des Netzes gesunken ist. War der Lernschritt erfolgreich, werden die neu berechneten Gewichte übernommen.

Im Allgemeinen wird zum Training eines KNN ein Musterdatensatz aus einer endlichen Anzahl von Eingangs-/Ausgangsdatenpaarungen benötigt. Bei in Echtzeit laufenden Anwendungen steht ein solcher Datensatz jedoch nicht zur Verfügung und so wurden in der Literatur bisher grundsätzlich zwei Lösungsvorschläge präsentiert. Eine vorgeschlagene Methode ist das ständige Trainieren des Netzes lediglich auf Basis des gerade aktuellen Netz-Ein- und -Ausganges. Hier wird der Lernerfolg nicht überprüft und somit ist nicht gewährleistet, dass der Ausgangsfehler in jedem Lernschritt sinkt. Ein mögliches Abbruchkriterium für das Lernen kann die Vorgabe eines minimalen Gradienten oder eines maximal tolerierten Fehlers sein. Die zweite Möglichkeit ist das so genannte Batch-Lernen. Bei dieser Methode wird während der Laufzeit ein Ringspeicher mit den letzten X Eingangs- und Ausgabewerten vorgehalten und die Anzahl der abgespeicherten Datenpaarungen entspricht der so genannten Batchsi-

ze. Dieser Datensatz wird nach einem Lernschritt zur Überprüfung des Lernerfolges genutzt und somit ist das Batch-Lernen die zu bevorzugende Methode. Aus diesem Grund wurde zu Beginn der Untersuchungen ein überwachtes Batch-Lernen implementiert. Bei der Erprobung musste aber festgestellt werden, dass das Verfahren auf der eingesetzten Echtzeithardware für das verwendete Netz nicht realisierbar ist. Zur Lösung des Problems wurde daraufhin die Batchsize zu eins gewählt und damit das adaptive KNN erfolgreich getestet.

Darüber hinaus muss an dieser Stelle angemerkt werden, dass aufgrund der hohen Tastrate analog zum ADNLRX-Modell eine Schrittweite α_{NN} zwischen den verwendeten Positionsmesswerten eingeführt wird. Somit ist es möglich, die Dynamik des Systems über einen größeren Zeithorizont abzudecken.

4.4 Störbeobachtung

Die letzte noch nicht behandelte Komponente der Trajektorienfolgeregelung aus Abbildung 4.1 ist der Störbeobachter. Hierbei fiel die Wahl auf ein Kalman Filter. Der Algorithmus sowie die Modellierung des Störbeobachters soll in den nächsten Abschnitten erläutert werden.

Bei einer Vielzahl von mechatronischen Systemen ist es nicht möglich, die inneren Systemzustände durch Messung zu bestimmen. Werden diese Größen aber für eine modellbasierte Regelung benötigt, müssen sie auf andere Weise bereitgestellt werden. Eine Möglichkeit ist die modellbasierte Schätzung der entsprechenden Zustände mittels eines Zustandsbeobachters. Dabei wird versucht, den Zustandsvektor mithilfe der gemessenen Systemausgänge zu berechnen. Ein speziell für stochastisch gestörte Signale entwickelter Algorithmus ist das so genannte Kalman Filter, welches 1960 von Rudolf Kalman vorgestellt wurde [66]. Für lineare, dynamische und stochastisch gestörte¹⁹ Systeme stellt der Algorithmus von Kalman einen optimalen Zustandsbeobachter dar. Darüber hinaus ist dieser durch die rekursive Berechnung für die Anwendung in Echtzeitsystemen prädestiniert.

In den Arbeiten von Ray *et al.* wird das Kalman Filter nicht als Zustandsbeobachter, sondern als Störbeobachter eingesetzt, der nicht messbare Störungen in reibbehafteten kinematischen Systemen schätzt [96, 97, 98]. Auf Basis dieser Schätzung werden diese Störeinflüsse vorsteuernd kompensiert und dadurch auftretende Schleppfehler deutlich reduziert. Besonders im Hinblick auf NPM-Systeme bietet sich der Einsatz eines solchen Schätzers an, denn unberücksichtigte Systemdynamiken und/oder stochastische exogene Störquellen üben einen signifikanten Einfluss auf das dynamische Systemverhalten aus.

Im folgenden Abschnitt wird zunächst die prinzipielle Funktionsweise eines Kalman Filters erläutert, während im zweiten Abschnitt die konkrete Anwendung zur Störgrößenkompensation vorgestellt wird [8, 141].

¹⁹ Hierbei wird davon ausgegangen, dass die stochastische Störung normalverteilt ist.

4.4.1 Prinzipielle Funktionsweise eines Kalman Filters

Das Kalman Filter basiert auf einer diskreten Zustandsraumdarstellung des betrachteten mechatronischen Systems:

$$\mathbf{x}(k) = \mathbf{A}_d \cdot \mathbf{x}(k-1) + \mathbf{B}_d \cdot \mathbf{u}(k) + \mathbf{v}(k) \quad (4.177)$$

$$\mathbf{y}(k) = \mathbf{C}_d \cdot \mathbf{x}(k) + \mathbf{o}(k) \quad (4.178)$$

Dabei werden das Prozessrauschen $\mathbf{v}(k)$ und das Messrauschen $\mathbf{o}(k)$ als mittelwertfreie, weiße gaußverteilte Rauschprozesse angenommen. Wenn die Kovarianzmatrizen dieser Rauschprozesse

$$\mathbf{Q}(k) = E[\mathbf{v}(k) \cdot \mathbf{v}^T(k)] \quad (4.179)$$

$$\mathbf{R}(k) = E[\mathbf{o}(k) \cdot \mathbf{o}^T(k)] \quad (4.180)$$

bekannt und $\mathbf{x}(k)$, $\mathbf{v}(k)$ und $\mathbf{o}(k)$ stochastisch unabhängige Größen sind, berechnet das Kalman Filter die optimale Schätzung $\hat{\mathbf{x}}(k)$ des Zustandsvektors $\mathbf{x}(k)$, sodass der mittlere quadratische Schätzfehler $E[(\mathbf{x}(k) - \hat{\mathbf{x}}(k)) \cdot (\mathbf{x}(k) - \hat{\mathbf{x}}(k))^T]$ minimal wird. Die Startwerte für den Zustandsvektor $\hat{\mathbf{x}}(0)$ und die Fehlerkovarianzmatrix $\mathbf{P}(0)$ ergeben sich aus dem Erwartungswert und der Varianz des Zustandes $\mathbf{x}(0)$:

$$\hat{\mathbf{x}}(0) = E[\mathbf{x}(0)] \quad (4.181)$$

$$\mathbf{P}(0) = \text{Var}(\mathbf{x}(0)) = E[(\mathbf{x}(0) - \hat{\mathbf{x}}(0)) \cdot (\mathbf{x}(0) - \hat{\mathbf{x}}(0))^T] \quad (4.182)$$

Da bei geringer Prozesskenntnis diese Größen a priori selten bekannt sind, haben sich in der praktischen Anwendung folgende Startwerte als geeignet erwiesen:

$$\hat{\mathbf{x}}(0) = \mathbf{0} \quad (4.183)$$

$$\mathbf{P}(0) = c \cdot \mathbf{I} \quad (4.184)$$

Je ungenauer das Vorwissen ist, umso höher wird der Parameter c gewählt, denn c ist ein Maß für die Anpassung an das Systemverhalten und liegt im Bereich von $10^2 \leq c \leq 10^{10}$.

Wurde der Algorithmus initialisiert, wechseln sich ein Prädiktionsschritt und ein Korrekturschritt gegenseitig ab.

Im Prädiktionsschritt werden sowohl der Zustandsvektor $\tilde{\mathbf{x}}(k)$ als auch die Fehlerkovarianzmatrix $\tilde{\mathbf{P}}(k)$ zum aktuellen Zeitpunkt wie folgt geschätzt:

$$\tilde{\mathbf{x}}(k) = \mathbf{A}_d \cdot \mathbf{x}(k-1) + \mathbf{B}_d \cdot \mathbf{u}(k) \quad (4.185)$$

$$\tilde{\mathbf{P}}(k) = \mathbf{A}_d \cdot \mathbf{P}(k-1) \cdot \mathbf{A}_d^T + \mathbf{Q}(k-1) \quad (4.186)$$

Im darauffolgenden Korrekturschritt wird zunächst die neue Verstärkungsmatrix $\mathbf{K}_K(k)$ des Kalman Filters berechnet. Durch sie wird es möglich, aktuelle Messwerte mit in die Schätzung einzubeziehen. Diese nachträgliche Korrektur führt dazu, dass im nächsten Zeitschritt

der Zustandsvektor präziser vorhergesagt werden kann. Des Weiteren findet im Korrekturschritt eine Aktualisierung der Fehlerkovarianzmatrix statt. Somit erfolgt im Korrekturschritt:

$$\mathbf{K}_k(k) = \tilde{\mathbf{P}}(k) \cdot \mathbf{C}_d^T \cdot [\mathbf{C}_d \cdot \tilde{\mathbf{P}}(k) \cdot \mathbf{C}_d^T + \mathbf{R}(k)]^{-1} \quad (4.187)$$

$$\hat{\mathbf{x}}(k) = \tilde{\mathbf{x}}(k) + \mathbf{K}_k(k) \cdot [\mathbf{y}(k) - \mathbf{C}_d \cdot \tilde{\mathbf{x}}(k)] \quad (4.188)$$

$$\mathbf{P}(k) = [\mathbf{I} - \mathbf{K}_k(k) \cdot \mathbf{C}_d] \cdot \tilde{\mathbf{P}}(k) \quad (4.189)$$

Entsprechend den Anfangswerten für $\mathbf{x}(k)$ und $\mathbf{P}(k)$ ist auch die Bestimmung der Kovarianzmatrizen $\mathbf{R}(k)$ und $\mathbf{Q}(k)$ der Rauschprozesse in der realen Anwendung ausgesprochen schwierig. Daher werden diese Matrizen beim Entwurf eines Kalman Filters oft als Tuningparameter angesehen und in Abhängigkeit des Systems experimentell ermittelt.

4.4.2 Störbeobachtung in NPM-Systemen

Wie bereits in [8] gezeigt wurde, können durch Messung der dem System zugeführten Kraft $F_A(t)$ und der resultierenden Position $x(t)$ unmodellerte Systemdynamiken sowie exogene Störeinflüsse durch ein Kalman Filter geschätzt werden. Das Ziel hierbei ist die Kompensation dieser Störungen, indem die geschätzte Kraft additiv auf die vom Regler berechnete Stellgröße aufgeschaltet wird (siehe Abbildung 4.18). Eine ähnliche Nutzung des Kalman Filters wurde bereits in den Arbeiten von Ramasubramaniam *et al.* vorgestellt und erfolgreich eingesetzt [96, 97]. Im Gegensatz zu Ramasubramaniam wird das hier vorgestellte Kalman Filter zusätzlich zu einer Vorsteuerung eingesetzt, und somit muss das Systemmodell aus Gleichung (3.1) erweitert werden. Es gilt für eine Bewegungsachse:

$$m \cdot \ddot{x}(t) + F_R(t) + F_S(t) = F_A(t) \quad (4.190)$$

Die dem System zugeführte Antriebskraft $F_A(t)$ beschleunigt die bewegte Masse m und wirkt dabei der Reibkraft $F_R(t)$ und der von den Störungen erzeugten Kraft $F_S(t)$ ²⁰ entgegen. Da die Reibkraft $F_R(t)$ durch ein adaptives Reibmodell vorgesteuert wird, muss diese Schätzung als zusätzliche Eingangsgröße mit im Zustandsraummodell des Kalman Filters berücksichtigt werden. Weiterhin wird die nicht messbare Störung als Zustand definiert, damit sie in jedem Tastschritt neu geschätzt werden kann. Somit ergibt sich die folgende Zustandsraumdarstellung:

$$\begin{bmatrix} \dot{x}(t) \\ \ddot{x}(t) \\ \dot{F}_S(t) \\ \ddot{F}_S(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{m} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \\ F_S(t) \\ \dot{F}_S(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{1}{m} & -\frac{1}{m} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} F_A(t) \\ F_R(t) \end{bmatrix} \quad (4.191)$$

²⁰ Diese unbekannte Kraft wird durch das Kalman Filter geschätzt und im darauffolgenden Tastschritt von der Trajektorienfolgeregelung verwendet (siehe Abbildung 4.1).

$$x(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(t) \\ \dot{x}(t) \\ F_s(t) \\ \dot{F}_s(t) \end{bmatrix}$$

Bei dieser Formulierung wird die Annahme getroffen, dass die Störung in jedem Abtastintervall eine lineare Funktion der Zeit ist. Dies gilt jedoch nicht für den gesamten zeitlichen Verlauf, denn im Korrekturschritt des Kalman Filters wird unabhängig von der Modellierung die Störgröße in Richtung des tatsächlichen Wertes korrigiert. Somit schränkt die in Gleichung (4.191) getroffene Annahme den zeitlichen Verlauf der Störgröße nicht ein und es kann in Anbetracht der hohen Tastzeit ein beliebiges Verhalten nachgebildet werden.

Für die Anwendung des Algorithmus von Kalman muss die kontinuierliche Zustandsraumdarstellung diskretisiert werden. Bei Verwendung eines Haltegliedes 0-ter Ordnung und der Tastzeit T können die diskretisierten Systemmatrizen berechnet werden mit [83]:

$$\mathbf{A}_d = e^{\mathbf{A} \cdot T} \quad (4.192)$$

$$\mathbf{B}_d = \int_0^T e^{\mathbf{A} \cdot t} dt \cdot \mathbf{B} \quad (4.193)$$

$$\mathbf{C}_d = \mathbf{C} \quad (4.194)$$

Aus diesen Gleichungen resultiert die folgende diskrete Zustandsraumdarstellung des Systems:

$$\begin{bmatrix} x(k) \\ \dot{x}(k) \\ F_s(k) \\ \dot{F}_s(k) \end{bmatrix} = \begin{bmatrix} 0 & T & -\frac{T^2}{2 \cdot m} & -\frac{T^3}{6 \cdot m} \\ 0 & 0 & -\frac{T}{m} & -\frac{T^2}{2 \cdot m} \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x(k-1) \\ \dot{x}(k-1) \\ F_s(k-1) \\ \dot{F}_s(k-1) \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2 \cdot m} & -\frac{T^2}{2 \cdot m} \\ \frac{T}{m} & -\frac{T}{m} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} F_A(k) \\ F_R(k) \end{bmatrix} \quad (4.195)$$

$$x(k) = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x(k) \\ \dot{x}(k) \\ F_s(k) \\ \dot{F}_s(k) \end{bmatrix}$$

Das diskrete Zustandsraummodell aus Gleichung (4.195) ist nun die Grundlage für den in Abschnitt 4.4.1 vorgestellten Algorithmus von Kalman.

5 Leistungsfähigkeit des dynamischen Regelungskonzeptes

Im vorangegangenen Kapitel wurde ein modellbasiertes Regelungskonzept für NPM-Systeme vorgestellt. Dieser aus der Robotik stammende Regelungsansatz ermöglicht eine deutliche Steigerung der erzielbaren Systemdynamik bei gleichbleibender Regelgüte. Der Kern des Konzeptes sind mathematische Modelle, die eine vorsteuernde Kompensation des Systemverhaltens gestatten. So kann eine signifikante Dynamiksteigerung erreicht werden, denn dieser hybride Ansatz vereinigt die Vorteile einer Steuerung mit den Vorteilen einer Regelung.

Prinzipiell besteht eine solche Trajektorienfolgeregelung aus den Regelungskomponenten Bahnplanung, Vorsteuerung, Störbeobachtung und unterlagerte Regelung. Sowohl die Struktur als auch die Funktionsweise aller beteiligten Regelungskomponenten wurden bereits ausführlich in Kapitel 4 dargelegt. Mithilfe einer Leistungsbewertung soll in diesem Kapitel deren Funktionstüchtigkeit auf Nanometerebene gezeigt werden.

Zu Beginn werden die durchgeführten Experimente eingehend beschrieben und die Auswahl der eingesetzten Trajektorien begründet. Anschließend wird anhand von experimentellen Daten der Einfluss jeder einzelnen Regelungskomponente auf die Güte des geregelten Systems analysiert und diskutiert. Abgeschlossen wird das Kapitel mit einer ausführlichen Bewertung der Ergebnisse.

5.1 Untersuchte Trajektorien

Wie bereits die Parameteridentifikation aus Abschnitt 3.4.3 wird auch die Leistungsbewertung des entworfenen Regelungssystems anhand der Y-Achse durchgeführt. Hierzu werden zwei Experimente durchgeführt. Als Referenztrajektorie dient im ersten Versuch eine periodische Bewegung mit einer Amplitude von 10.000 nm und einer Frequenz von 1 Hz. Bei dieser Bewegung durchläuft das tribologische System permanent alle Reibdomänen und somit ist dieser Versuch prädestiniert für die Bewertung der Leistungsfähigkeit des Regelungssystems im Hinblick auf die Kompensation der dominanten Reibeffekte. In einem zweiten Experiment wird die Y-Achse über eine Strecke von 10 mm mit einer relativ hohen Maximalgeschwindigkeit von 1 mm/s verfahren. Die kinematischen Solltrajektorien zu dieser linearen Bewegung sind in Abbildung 5.1 dargestellt. Wie zu sehen ist, schließt die hier abgebildete lineare Bewegung auch die Rückfahrt mit ein. Deren kinematisches Profil ist identisch mit dem der Hinfahrt. Das Experiment ermöglicht die Einschätzung der Leistungsfähigkeit des Regelungssystems bei relativ hohen Geschwindigkeiten.

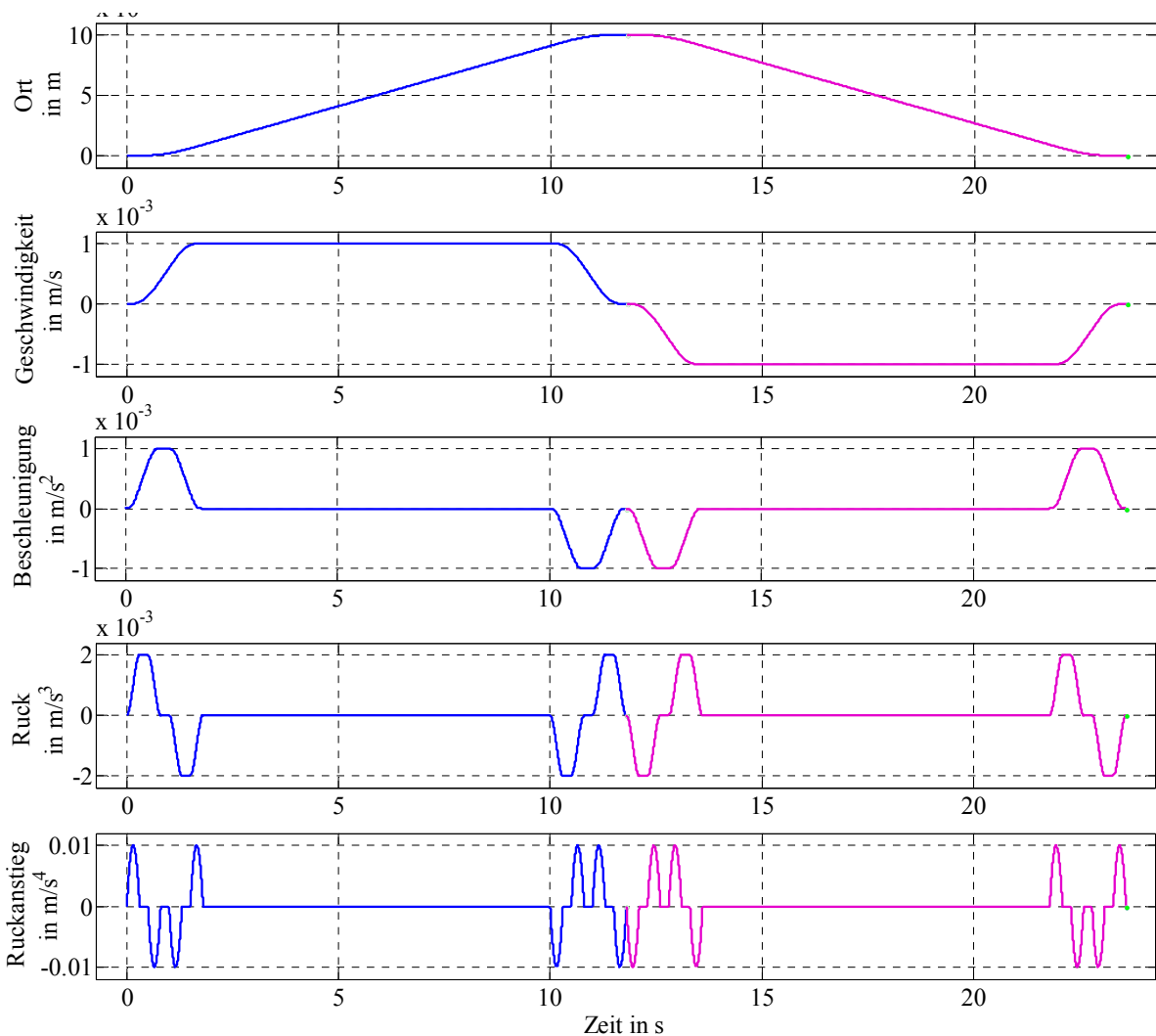


Abbildung 5.1: Sollwerte aller kinematischen Größen für eine lineare Bewegung

Darüber hinaus kann mit diesem Versuch die Reaktion des Regelungssystems auf stochastische Störungen wie beispielweise Verschmutzungen oder Produktionstoleranzen in den Führungen, Elastizität der Flexbandkabel, die Position der Kugelkäfige, et cetera untersucht werden.

5.2 Einfluss der Regelungskomponenten auf die erzielbare Regelgüte

Zur Bewertung der Regelgüte wird wie in den vorangegangenen Kapiteln der mittlere quadratische Fehler herangezogen. Dieser ergibt sich aus dem dynamischen Regelfehler und ist definiert als:

$$\text{RMSE} = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N (x_{\text{soll}}(k) - x_{\text{mess}}(k))^2} \quad (5.1)$$

In Gleichung (5.1) repräsentiert N die Anzahl der Wertepaare im betrachteten Datensatz, $x_{\text{soll}}(k)$ die Sollposition und $x_{\text{mess}}(k)$ die tatsächliche Position.

Mit dem Ziel, die Leistungsfähigkeit der einzelnen Regelungskomponenten zu untersuchen, wird zu Beginn der experimentellen Verifikation die Y-Achse des Demonstrators mit einer sinusförmigen Solltrajektorie beaufschlagt. Wird das System lediglich mit einem PID-Regler oder einem PI-Zustandsregler betrieben, führt dies zu großen Schleppfehlern. Weiterhin ist sowohl in Abbildung A 22 als auch in Abbildung A 23 (siehe Anhang C) ein deutlicher Phasenverzug erkennbar. Der maximale Regelfehler liegt bei circa 500 nm. Entsprechend groß wird auch der RMSE mit 390,99 nm beziehungsweise 290,93 nm.

Eine deutliche Verbesserung ergibt sich durch die Vorsteuerung des inversen Systemverhaltens. Wie Tabelle 5.1 zeigt, gelingt dies mit beiden adaptiven Reibmodellen nahezu gleich gut. In Abbildung A 24 bis Abbildung A 27 ist der dynamische Regelfehler für alle Kombinationen aus Vorsteuerung und unterlagerter Regelung dargestellt. Der Phasenverzug im Regelfehler ist nicht mehr sichtbar und es kann dadurch der RMSE deutlich reduziert werden. Der maximale Fehler liegt klar über der erzielten mittleren Bahnabweichung und tritt kurz nach einer Richtungsumkehr auf.

Durch den Einsatz des entworfenen Störbeobachters wird der maximale Fehler sowie der RMSE weiter verkleinert (siehe Abbildung A 28 bis Abbildung A 31). Die Ergebnisse für alle Kombinationen aus Vorsteuerung, unterlagerter Regelung und Störbeobachtung sind in Tabelle 5.1 zusammenfassend dargestellt.

Tabelle 5.1: Mittlerer quadratischer Bahnfehler bei einer sinusförmigen Bewegung unter Verwendung der untersuchten Reglerkonfigurationen

	Nur Regelung	ADNLRX	KNN	ADNLRX mit Störbeobachter	KNN mit Störbeobachter
PID-Regler	390,99 nm	31,44 nm	63,9 nm	16,51 nm	59,76 nm
PI- Zustandsregler	290,93 nm	30,89 nm	54,87 nm	14,32 nm	30,05 nm

Resümierend kann festgehalten werden, dass eine modellbasierte Regelung aus PID-Regler, ADNLRX-Modell und Störbeobachter in der Lage ist, bei einer sinusförmigen Bewegung den RMSE um Faktor 23 und den maximalen Fehler um Faktor sieben zu verkleinern.

Um das Verhalten des Regelungssystems bei hohen Geschwindigkeiten zu analysieren, wird eine hochdynamische Bewegung über eine Strecke von 10 mm ausgeführt. In Analogie zum vorangegangenen Experiment entstehen große Schleppfehler, wenn nur die unterlagerte Regelung aktiv ist. Besonders zu Beginn der Bewegung zeigen beide eingesetzten Reglertypen signifikante Schwächen und der Regelfehler beträgt hier nahezu 1000 nm (Abbildung A 32 und Abbildung A 33).

Der Einsatz einer Vorsteuerung verbessert das dynamische Verhalten des Regelungssystems bei hohen Geschwindigkeiten ausgesprochen deutlich (siehe Abbildung A 34 bis Abbildung

A 37). Bei diesem Szenario ist eine Kombination aus Zustandsregler und KNN allen anderen Konfigurationen aus unterlagerter Regelung und Vorsteuerung überlegen.

Wird zusätzlich zur Vorsteuerung ein Störbeobachter eingesetzt, verschieben sich die Ergebnisse leicht zu Gunsten des ADNLRX-Modells. Anhand von Abbildung A 38 bis Abbildung A 41 wird besonders deutlich, dass der Störbeobachter in der Lage ist, die großen Regelabweichungen während der Anfahrt zu beseitigen. Alle Ergebnisse sind zur besseren Vergleichbarkeit noch einmal in Tabelle 5.2 zusammengefasst.

Tabelle 5.2: Mittlerer quadratischer Bahnfehler bei einer linearen Bewegung unter Verwendung der untersuchten Reglerkonfigurationen

	Nur Regelung	ADNLRX	KNN	ADNLRX mit Störbeobachter	KNN mit Störbeobachter
PID-Regler	315,80 nm	97,58 nm	39,68 nm	14,53 nm	22,77 nm
PI- Zustandsregler	288,27 nm	52,16 nm	44,18 nm	14,81 nm	15,87 nm

5.3 Bewertung der Ergebnisse

Bei der Auswahl eines geeigneten Regelungssystems müssen neben der Regelgüte natürlich auch Beschränkungen des mechatronischen Systems mit berücksichtigt werden. Besonders die Rechenleistung des verwendeten Echtzeitsystems setzt den implementierbaren Regelalgorithmen Grenzen. Um diesem Umstand Rechnung zu tragen, werden die durchschnittlichen Laufzeiten aller möglichen Reglerkonfigurationen untersucht. Anhand der Grafik in Abbildung 5.2 ist klar zu erkennen, dass das KNN im Vergleich zu den übrigen Reglerkomponenten einen signifikanten Einfluss auf die Laufzeit des gesamten Regelalgorithmus nimmt. Diese Beobachtung kann damit erklärt werden, dass die Anzahl der zu schätzenden Parameter des KNN um ein Vielfaches größer ist als die des ADNLRX-Modells. Da die erzielbare Vorhersagegüte bei beiden Modellen nahezu gleich ist, kann das ADNLRX-Modell als deutlich effizienter angesehen werden.

Bezüglich der verbleibenden Regelungskomponenten ist eine derart klare Aussage nicht möglich. Weder der PID-Regler noch der PI-Zustandsregler oder der Störbeobachter unterscheiden sich signifikant in der benötigten Rechenzeit. Aufgrund seiner meist besseren Leistungen wird daher der PI-Zustandsregler mit Führungsgrößenaufschaltung als unterlagerter Regler favorisiert. In gleicher Weise ist der Einsatz eines Störbeobachters nach den Ergebnissen der vorliegenden Arbeit unabdingbar, denn eine effektive Störbeobachtung trägt nochmals deutlich zur Reduktion des Schleppfehlers bei. Weiterhin zeigen die Untersuchungsergebnisse, dass eine Vorsteuerung der inversen Systemdynamik die Regelgüte am signifikantesten beeinflusst. Dieser Sachverhalt wirkt sich natürlich auch auf den Einfluss der unterlagerten Regelung aus, denn deren Leistungsfähigkeit spielt nur noch eine untergeordnete Rolle.

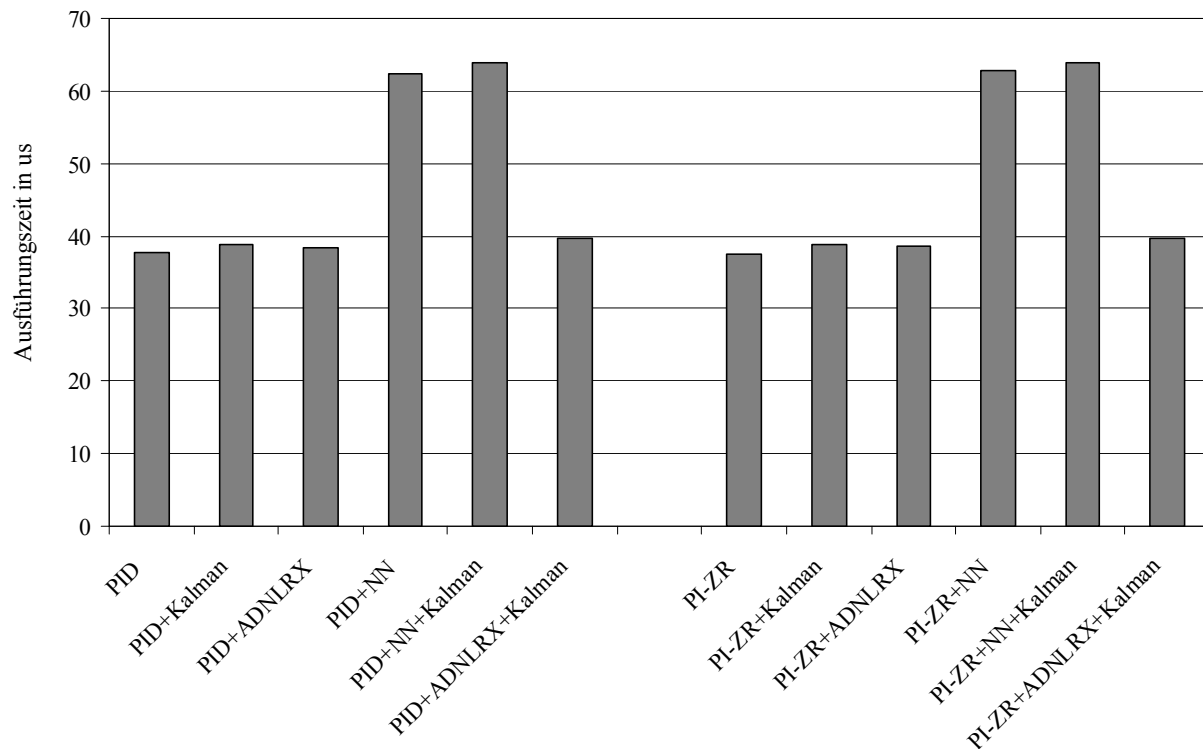


Abbildung 5.2: Gegenüberstellung der Ausführungszeiten bei Verwendung der untersuchten Reglerkonfigurationen

Auf Grundlage dieser Untersuchungen kann festgehalten werden, dass eine Trajektorienfolge-regelung aus ADNLRX-Modell, PI-Zustandsregler mit Führungsgrößenaufschaltung und einem Störbeobachter auf Kalman-Filter-Basis den besten Kompromiss aus Regelgüte und benötigter Rechenzeit darstellt.

Darüber hinaus sollte an dieser Stelle auf ein grundsätzliches Problem bei der Regelung von NPM-Systemen hingewiesen werden. Wie in Abschnitt 3.3 gezeigt wurde, besitzt das Antriebssystem eine Eckfrequenz von 10 kHz. Dieser Umstand begrenzt folglich auch die maximale Geschwindigkeit des Regelungssystems, denn eine schnellere Berechnung der Stelleingriffe kann nicht umgesetzt werden. Daher wurde die Tastfrequenz des Regelungssystems ebenfalls zu 10 kHz gewählt und es erfolgt nur alle 100 μs ein Stelleingriff. Somit kann in Abhängigkeit der Geschwindigkeit eine „räumliche“ Schrittweite berechnet werden, die angibt nach welcher zurückgelegten Strecke ein neuer Stelleingriff erfolgt (siehe Abbildung 5.3). In Anbetracht der Tatsache, dass die Arbeitsgeschwindigkeiten von NPM-Systemen einen Bereich zwischen 10 nm/s bis 30.000.000 nm/s umfassen, limitiert die Zeit zwischen zwei Stelleingriffen besonders bei hohen Geschwindigkeiten natürlich die erreichbare Regelgüte. Mit dem Ziel, diesen Zusammenhang zu zeigen, wurde die in Abschnitt 5.1 vorgestellte lineare Bewegung mit verschiedenen maximalen Geschwindigkeiten ausgeführt und dabei der RMSE ausgewertet. Zur Regelung der Y-Achse kam die favorisierte Reglerkonfiguration aus PI-Zustandsregler, ADNLRX-Modell und Störbeobachter zum Einsatz. Die Ergebnisse der Analyse sind in Abbildung 5.3 dargestellt. Die Grafik zeigt klar, dass der RMSE mit der steigenden Geschwindigkeit wächst. Nichtsdestotrotz ist mit der eingesetzten Trajektorienfolge-

regelung ein RMSE erreichbar, der merklich unterhalb der „räumlichen“ Schrittweite liegt. Dieses Verhältnis zeigt deutlich die Güte der eingesetzten modellbasierten Regelung.

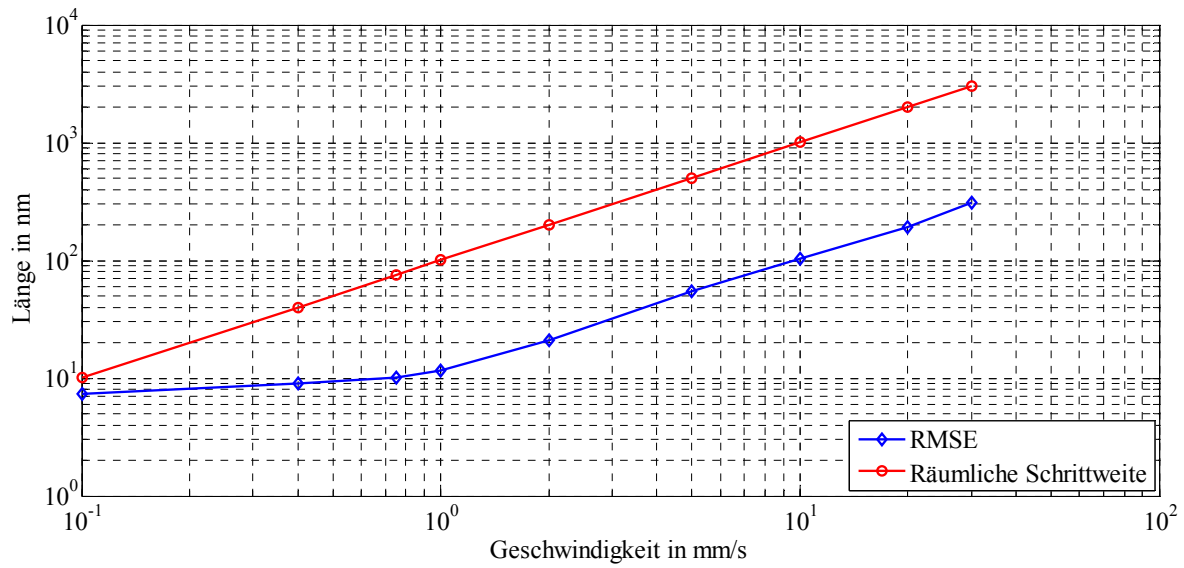


Abbildung 5.3: Entwicklung des RMSE in Abhängigkeit der Geschwindigkeit

6 Regelung von X- und Y- Achse

In Kapitel 5 konnte gezeigt werden, dass mithilfe des entwickelten Regelungskonzeptes der dynamische Regelfehler von NPM-Systemen signifikant verringert werden kann. Bisher wurden alle Untersuchungen lediglich anhand der Y-Achse des xy-Demonstrators durchgeführt und somit stellt sich die Frage, wie dieses Regelungskonzept auf beide Bewegungsachsen übertragen werden kann. Als Lösung wird in diesem Kapitel das Konzept einer dezentralen Trajektorienfolgeregelung vorgeschlagen und deren Funktionstüchtigkeit anhand ausgewählter experimenteller Untersuchungen nachgewiesen.

6.1 Dezentrales Regelungskonzept

Aufgrund der Tatsache, dass der xy-Demonstrator ein Mehrgrößensystem ist, muss beim Regelungsentwurf das Gesamtsystem betrachtet werden. Im Fokus stehen hierbei Verkopplungen zwischen den zu regelnden Teilsystemen, denn wenn diese unberücksichtigt bleiben, kann das Gesamtsystem in ungünstigen Situationen zur Instabilität neigen.

Da die verwendeten Regelalgorithmen auf den Messwerten der Interferometer beruhen und deren gemeinsamer Bezugspunkt die Spiegelecke ist, stellt diese die stärkste Verkopplung zwischen den beiden Bewegungsachsen dar. Im Falle einer Verdrehung der Spiegelecke um die Z-Achse kommt es zu Kosinusfehlern bei der Positionsbestimmung [37]. Mit dem Ziel, diese gegenseitige Beeinflussung zu kompensieren, werden die Positionsmesswerte auf Basis des einmalig bestimmten Rotationswinkels korrigiert. Neben dem Messprinzip beeinflussen sich die Achsen des xy-Demonstrators auch aus konstruktiven Gründen (siehe Abschnitt 2.1). Bedingt durch die Doppel-H-Struktur ist der X-Slider innerhalb des Y-Sliders angebracht und es wirken alle Momente der X-Achse direkt auf die Y-Achse. Diese Verkopplungen können aber aus regelungstechnischer Sicht als vernachlässigbar klein angesehen werden, denn die kinematischen Größenordnungen der Bewegungen des xy-Demonstrators bewegen sich im einstelligen Zentimeterbereich.

Somit existieren nur leichte Verkopplungen zwischen den Teilsystemen und es wird zur Regelung des Gesamtsystems ein dezentraler Regelungsansatz vorgeschlagen. Wie in Abbildung 6.1 zu sehen ist, wird jede Achse des Demonstrators von einem lokalen Trajektorienfolgeregler betrieben und lediglich der Bahnplanungsalgorithmus generiert die kinematischen Sollwerte für alle lokalen Regler. Die Synchronisation der Einzelachsregelungen garantiert dabei der Bahnplaner. Dieser Algorithmus generiert unter Kenntnis der elektrischen und mechanischen Leistungsfähigkeit des Systems die Sollwerttrajektorien für jeden Einzelachsregler.

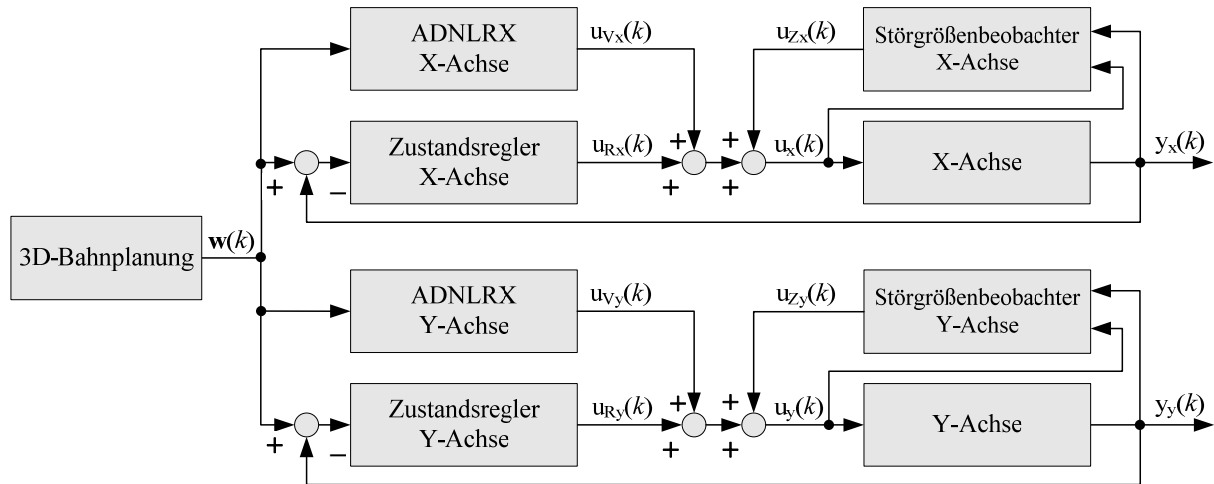


Abbildung 6.1: Dezentrales modellbasiertes Regelungskonzept für den xy-Demonstrator

Wie bereits in Kapitel 4.1 ausführlich beschrieben wurde, generiert der Algorithmus unter Kenntnis der elektrischen und mechanischen Leistungsfähigkeit des Systems die Sollwerttrajektorien für jeden Einzelachsregler und eine Superposition dieser Teilbewegungen ergibt die geplante Trajektorie der Tool Center Points in der xy-Ebene²¹.

6.2 Experimentelle Untersuchungen

Um die Funktionstüchtigkeit des vorgeschlagenen dezentralen Regelungskonzeptes zu zeigen, werden zwei Arten von Experimenten durchgeführt. Zum einen handelt es sich hierbei um Kreise mit unterschiedlichen Radien und zum anderen um komplexe zweidimensionale Figuren bestehend aus Geraden, Kreissegmenten und Bloss-Kurven. Bei diesen Experimenten wird das vorgeschlagene Konzept der dezentralen Trajektorienfolgeregelung mit einer dezentralen Regelung mittels klassischen PID-Reglern verglichen. Auch hier wird zur Bewertung der Leistungsfähigkeit die mittlere quadratische Länge des Bahnfehlervektors herangezogen. Das Fehlermaß kann wie folgt berechnet werden:

$$\text{RMSE} = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N \left[(x_{\text{mess}}(k) - x_{\text{soll}}(k))^2 + (y_{\text{mess}}(k) - y_{\text{soll}}(k))^2 \right]} \quad (6.1)$$

6.2.1 Kreise in der xy-Ebene

Mit dem Ziel, die Leistungsfähigkeit des dezentralen modellbasierten Regelungskonzeptes zu zeigen, werden die Achsen des xy-Demonstrators derart verfahren, sodass der TCP²² einen Kreisbogen von 720° durchfährt. Der Kreisbogen besitzt zunächst einen Durchmesser von 1 mm, da so das Verhalten des Regelungssystems bei makroskopischen Bewegungen charakterisiert werden kann. Die maximale Bahngeschwindigkeit beträgt bei diesem Experiment 300 µm/s.

²¹ Diese Einschränkung gilt nur in Anbetracht des eingesetzten zweidimensionalen Positioniersystems.

²² Bei diesen Experimenten wurde der Mittelpunkt der Spiegelecke als Tool Center Point definiert.

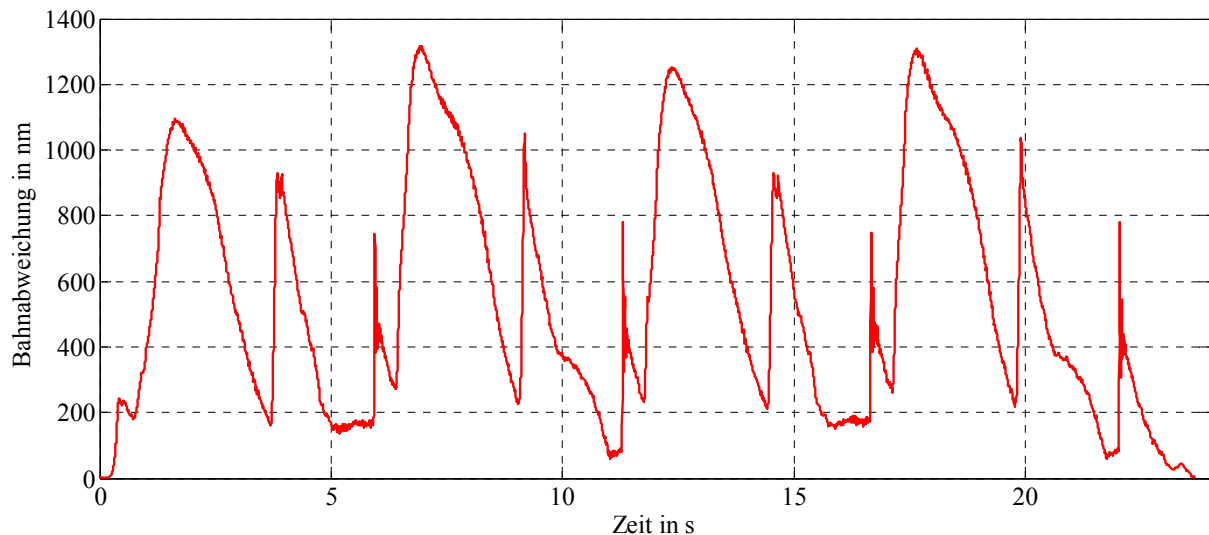


Abbildung 6.2: Bahnabweichung bei einer kreisförmigen Bewegung ($\varnothing = 1 \text{ mm}$, $v_{\max} = 0,3 \text{ mm/s}$) unter Nutzung eines dezentralen PID-Reglers

Werden als Achsregler experimentell parametrisierte PID-Regler verwendet, wird ein RMSE von 658,42 nm erreicht. Der zugehörige Bahnfehler ist in Abbildung 6.2 dargestellt. In dieser Grafik ist deutlich zu erkennen, dass die größten Bahnabweichungen von weit über 1000 nm an Kurvenpunkten auftreten, an denen eine der Achsen die Bewegungsrichtung umkehren muss. Wird im Vergleich dazu das vorgestellte modellbasierte Regelungskonzept eingesetzt, kann der RMSE um Faktor 14 auf 45,74 nm reduziert werden. Den Verlauf des Bahnfehlers bei Einsatz der Trajektorienfolgeregler zeigt Abbildung 6.3. Problematisch sind auch hier die Bahnorte, an denen eine Richtungsumkehr stattfindet (siehe Abbildung 6.4). Der maximale Fehler kann mit circa 550 nm aber mehr als halbiert werden. In Kapitel 5 konnte festgestellt werden, dass sich die erzielbare Regelgüte bei Einsatz der Trajektorienfolgeregeler indirekt proportional zur Geschwindigkeit verhält. Um diese Beziehung auch für zweidimensionale Bewegungen nachzuweisen, wird ein weiterer Kreisbogen von 720° abgefahren. Nun beträgt der Durchmesser aber nur $10 \text{ }\mu\text{m}$ und die Höchstgeschwindigkeit auf der Bahn nur $3 \text{ }\mu\text{m/s}$.

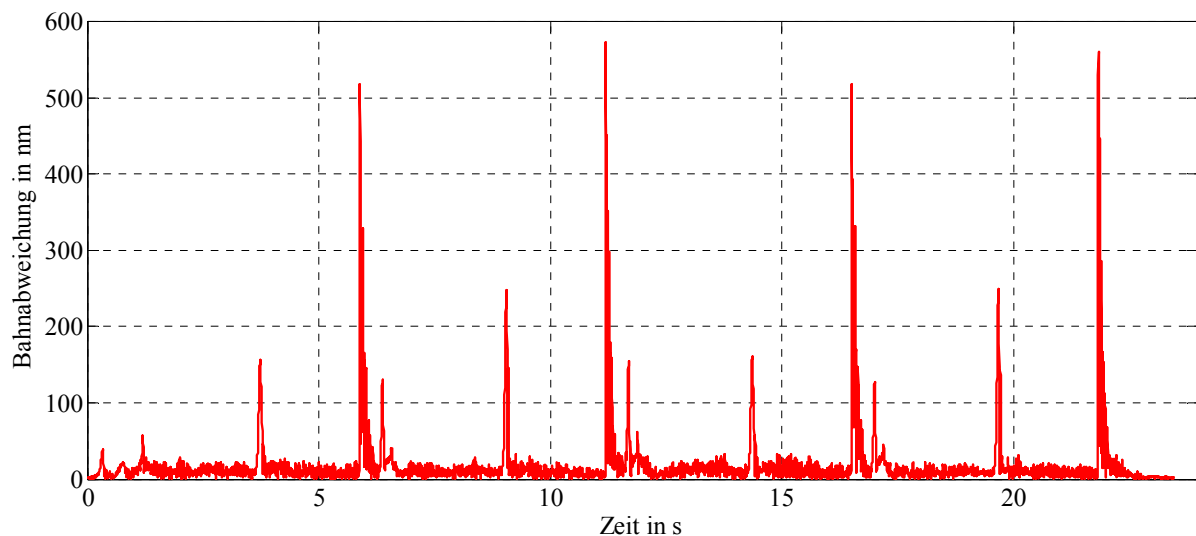


Abbildung 6.3: Bahnabweichung bei einer kreisförmigen Bewegung ($\varnothing = 1 \text{ mm}$, $v_{\max} = 0,3 \text{ mm/s}$) unter Nutzung des dezentralen modellbasierten Regelungskonzeptes

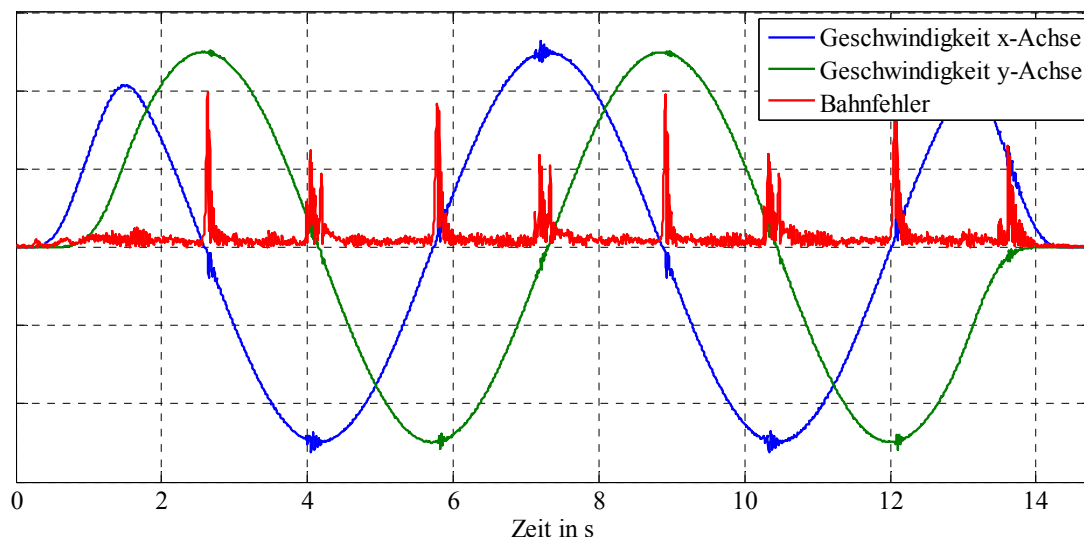


Abbildung 6.4: Zusammenhang zwischen dem Bahnfehler und der Geschwindigkeit²³ der Achsen

Abbildung 6.5 a) zeigt deutlich, dass die dezentralen PID-Achsregler nicht in der Lage sind, die Kontur des Kreises korrekt abzufahren.

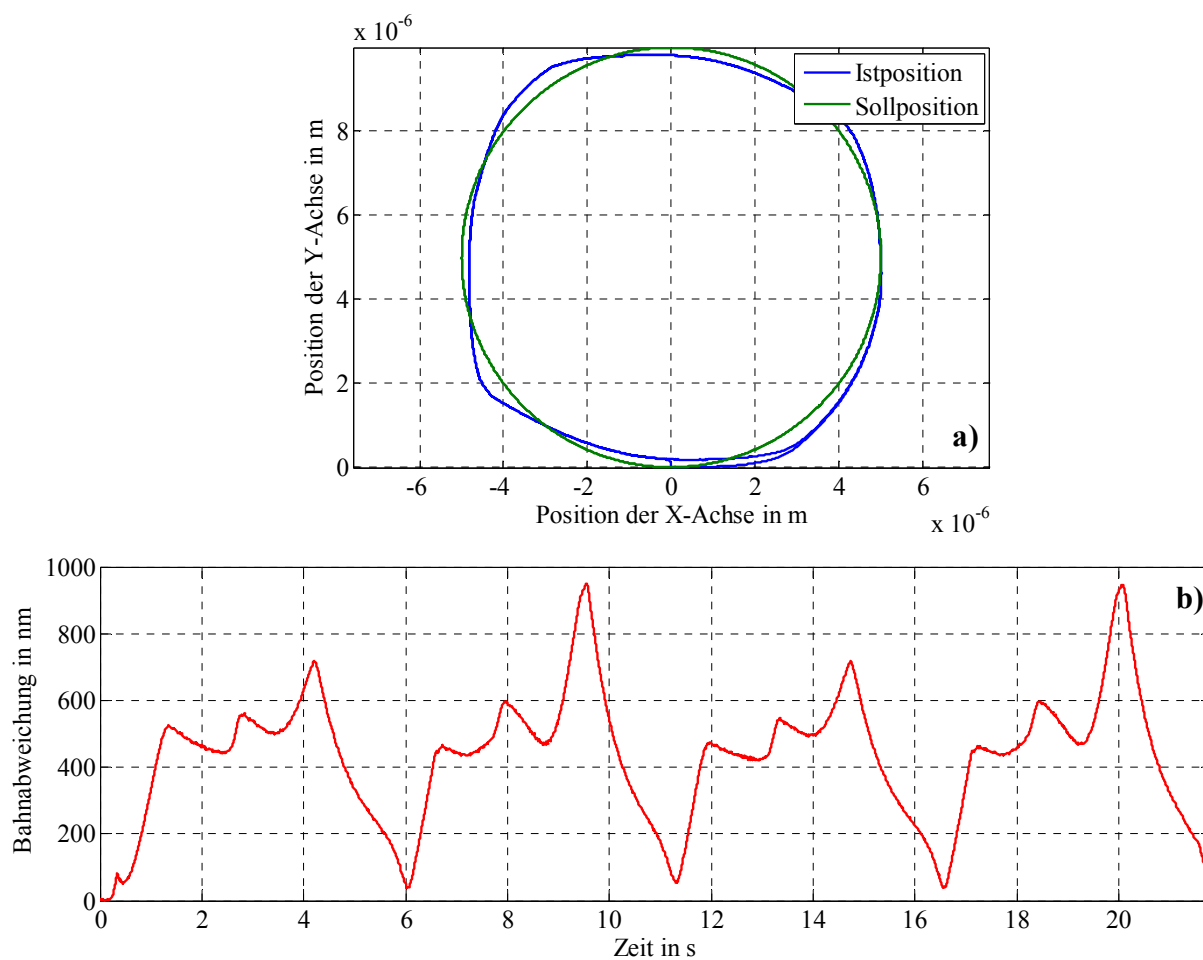


Abbildung 6.5: (a) Ist- und Sollbahn; (b) Bahnabweichung bei einer kreisförmigen Bewegung ($\varnothing = 10 \mu\text{m}$, $v_{\text{max}} = 3 \mu\text{m/s}$) unter Nutzung eines dezentralen PID-Reglers

²³ Die Amplitude der Geschwindigkeit wurde in dieser Grafik aus Gründen der Darstellbarkeit stark verkleinert.

In Analogie zum vorangegangenen Experiment ist der auftretende Bahnfehler an den Punkten der Richtungsumkehr am größten und beträgt hier nahezu 1000 nm (siehe Abbildung 6.5 b)). Der daraus resultierende RMSE ist mit 469,76 nm in Anbetracht des Kreisdurchmessers sehr groß. Werden anstelle der dezentralen PID-Regler die entworfenen Trajektorienfolgeregelungen verwendet, kann der TCP des xy-Demonstrators problemlos der Sollposition folgen (siehe Abbildung 6.6 a)). Der RMSE kann auf 10,19 nm gesenkt werden und in Spitze erreichen die Bahnabweichungen lediglich eine Größe von 34 nm (siehe Abbildung 6.6 b)). Bei diesem Experiment konnte somit der RMSE im Vergleich zu einer klassischen Regelung um den Faktor 46 und die maximale Bahnabweichung um Faktor 28 verringert werden. Diese Ergebnisse bestätigen eindeutig die Beobachtungen aus Kapitel 5.

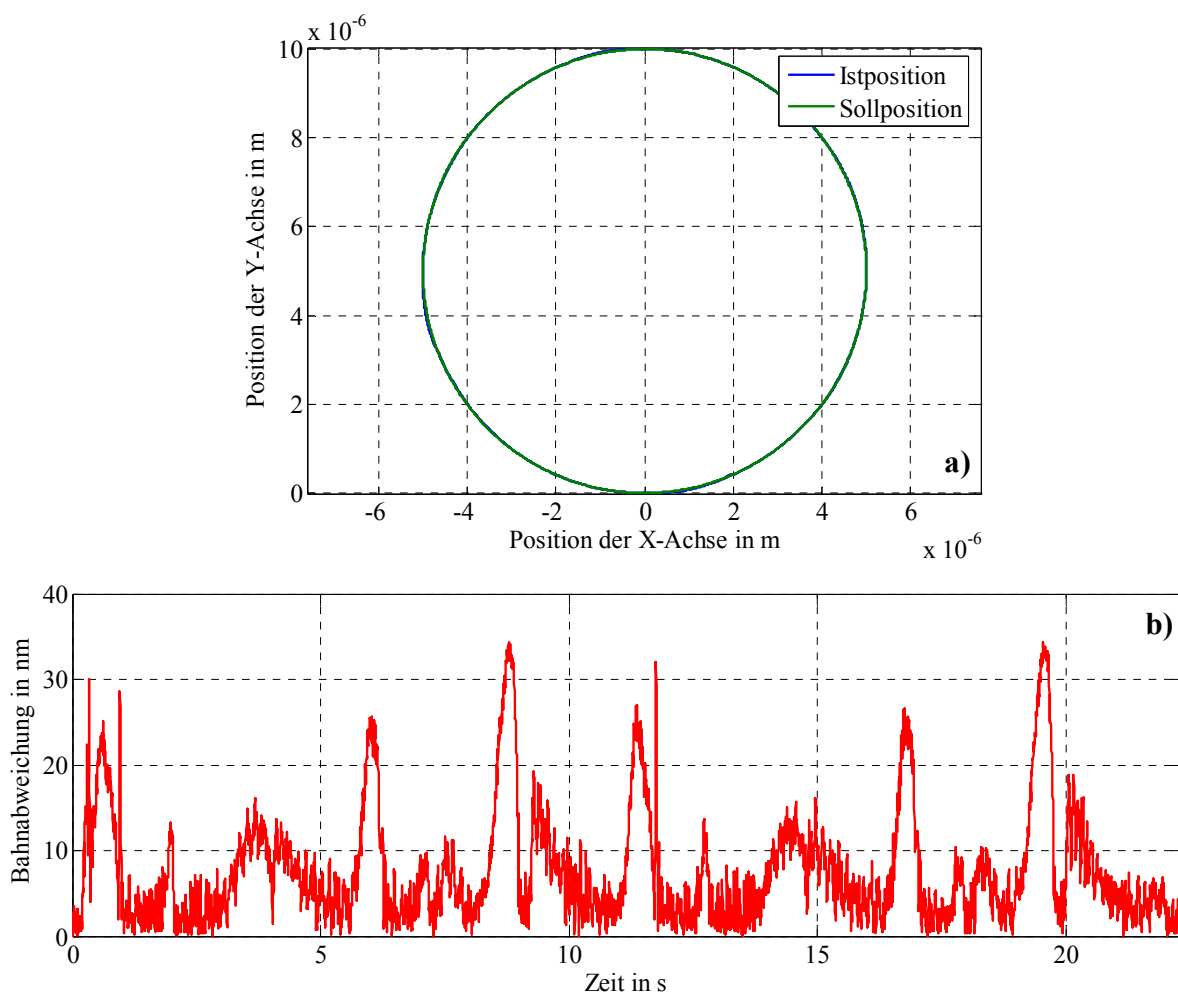


Abbildung 6.6: (a) Ist- und Sollbahn; (b) Bahnabweichung bei einer kreisförmigen Bewegung ($\varnothing = 10 \mu\text{m}$, $v_{\text{max}} = 3 \mu\text{m/s}$) unter Nutzung des dezentralen modellbasierten Regelungskonzeptes

6.2.2 Komplexe Figuren in der xy-Ebene

Zur weiteren Analyse werden neben den Kreisbögen zwei komplexe Figuren in der xy-Ebene abgefahren. Jede dieser Figuren wird in zwei Größen mit stark unterschiedlichen Bahngeschwindigkeiten ausgeführt und so ergeben sich vier Experimente. Dieses Vorgehen wurde gewählt, da nur so das Potential der dezentralen Trajektorienfolgeregelung an verschiedenen Arbeitspunkten untersucht werden kann.

Die erste Figur besteht aus zwei Kreisbögen und zwei sich schneidenden Geraden. Um eine kontinuierliche Durchfahrt zu ermöglichen, werden alle Wegsegmente mit Bloss-Kurven verschliffen. Wie in Abbildung A 42 (siehe Anhang D) zu sehen ist, besitzt die betrachtete Figur A das Aussehen einer Acht und dehnt sich über eine Fläche von $100 \times 40 \text{ mm}^2$ aus. Das Profil der Bahngeschwindigkeit ist in Abbildung A 43 gezeigt und beträgt im Durchschnitt $2,0 \text{ mm/s}$. Unter Nutzung der lokalen PID-Regler wird ein RMSE von $279,35 \text{ nm}$ erreicht. Der Verlauf des Bahnfehlers zeigt Abbildung 6.7. Im Gegensatz dazu belegt Abbildung 6.8 die Leistungsfähigkeit der modellbasierten Regelung. Der RMSE kann um den Faktor neun auf $31,01 \text{ nm}$ reduziert werden. Weiterhin ist im Vergleich der beiden Abbildungen deutlich erkennbar, dass die maximale Bahnabweichung auf etwa ein Viertel verkleinert werden kann.

Entsprechend den vorangegangenen Untersuchungen wirkt sich bei mikroskopischen Bewegungen die Umstellung des Regelungssystems noch deutlicher aus. Hierzu wurde Figur A um den Faktor 30 verkleinert und besitzt nunmehr eine Ausdehnung von circa $3,3 \times 1,3 \text{ mm}^2$ (siehe Abbildung A 44). Sie wird im Folgenden als Figur B bezeichnet. Entsprechend der Größenreduktion werden auch die Bahngeschwindigkeiten um Faktor 30 auf durchschnittlich $0,1 \text{ mm/s}$ gesenkt (siehe Abbildung A 45).

Abbildung 6.9 und Abbildung 6.10 zeigen den dynamischen Regelfehler ohne beziehungsweise mit dezentraler Trajektorienfolgeregelung. Der RMSE kann von $207,12 \text{ nm}$ auf $10,56 \text{ nm}$ vermindert werden, was einer Verkleinerung auf $5,1 \%$ entspricht. Die Reduzierung des maximalen Bahnfehlers um den Faktor vier bleibt nahezu gleich.

Mit der Intention, die bisherigen Beobachtungen an einem weiteren Experiment zu verifizieren, werden die identischen Untersuchungen an den Figuren aus Abbildung A 46 und Abbildung A 50 durchgeführt. Auch hier handelt es sich wieder um die gleiche Kurve, deren Größe lediglich variiert wurde. Die Kurven sind aus zwei Kreissegmenten und drei linearen Wegsegmenten zusammengesetzt, die von erweiterten Bloss-Kurven verschliffen wurden. Die makroskopische Figur C besitzt eine Ausdehnung von $60 \times 60 \text{ mm}^2$ und die durchschnittliche Bahngeschwindigkeit beträgt 3 mm/s .

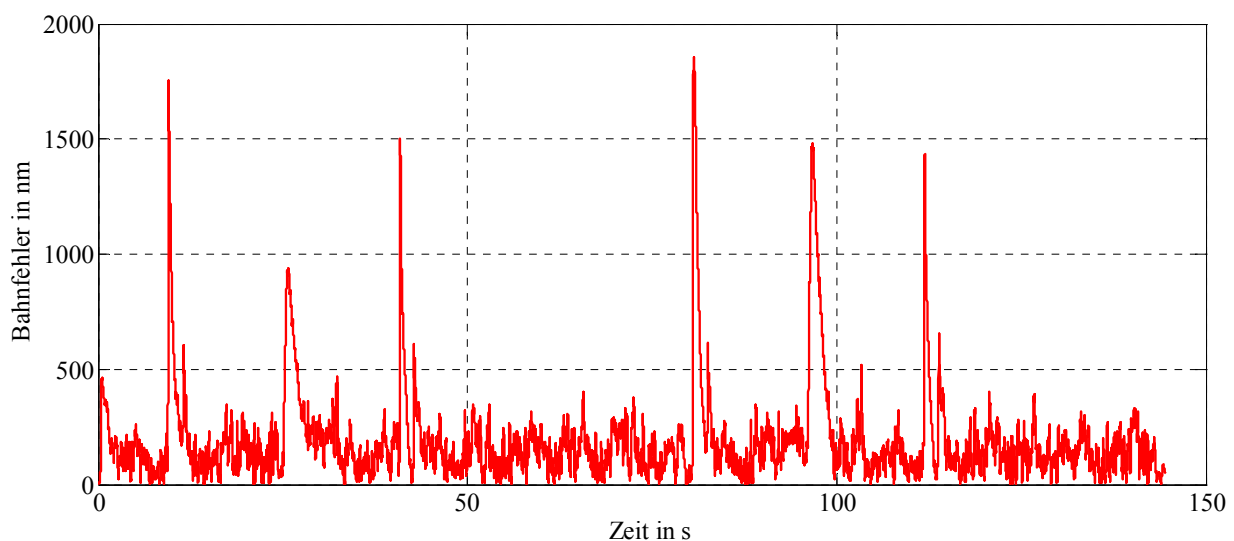


Abbildung 6.7: Bahnabweichung beim Abfahren von Figur A unter Nutzung eines dezentralen PID-Reglers

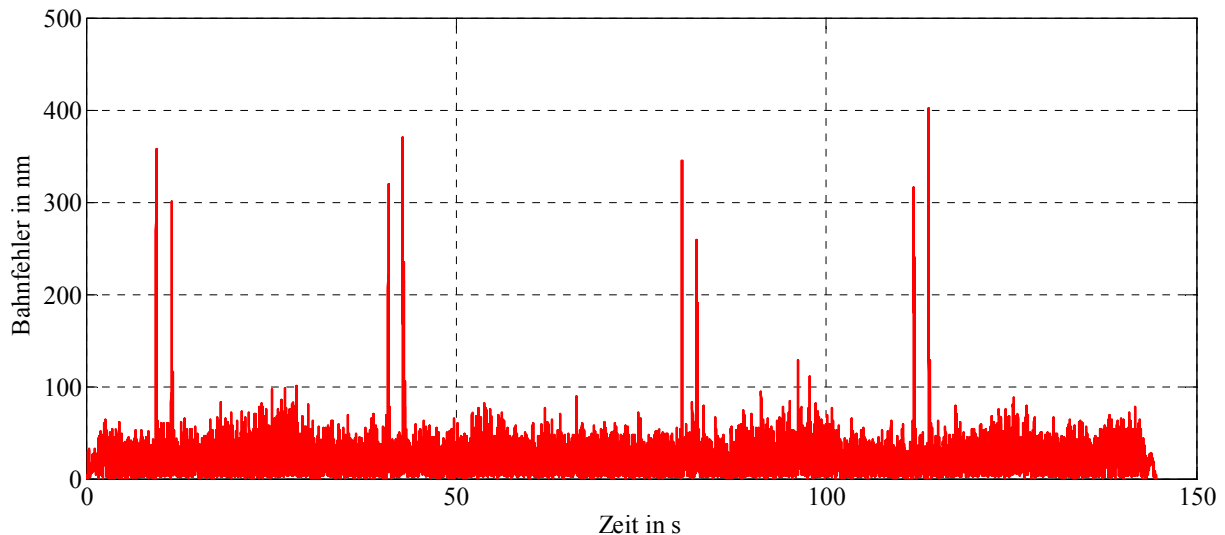


Abbildung 6.8: Bahnabweichung beim Abfahren von Figur A unter Nutzung des dezentralen modellbasierten Regelungskonzeptes

Bei Einsatz der klassischen Regelung ergibt sich ein RMSE von 347,57 nm und es entsteht an den Umkehrpunkten eine maximale Bahnabweichung von über 2000 nm. Die dezentrale Trajektorienfolgeregelung ist in der Lage, den RMSE um den Faktor sieben auf 49,94 nm zu reduzieren.

Bei Verkleinerung der Figur C um den Faktor 30 besitzt die nun entstandene Figur D nur noch eine Ausdehnung von $2 \times 2 \text{ mm}^2$ und der TCP bewegt sich mit durchschnittlich 0,1 mm/s. Anhand der Abbildung A 52 und Abbildung A 53 kann nachvollzogen werden, dass der RMSE beim Wechsel des Regelungskonzeptes von 207,85 nm auf 10,40 nm gesenkt werden kann. Im Gegensatz dazu wird die maximale Bahnabweichung aber nur um Faktor vier auf circa 260 nm verkleinert.

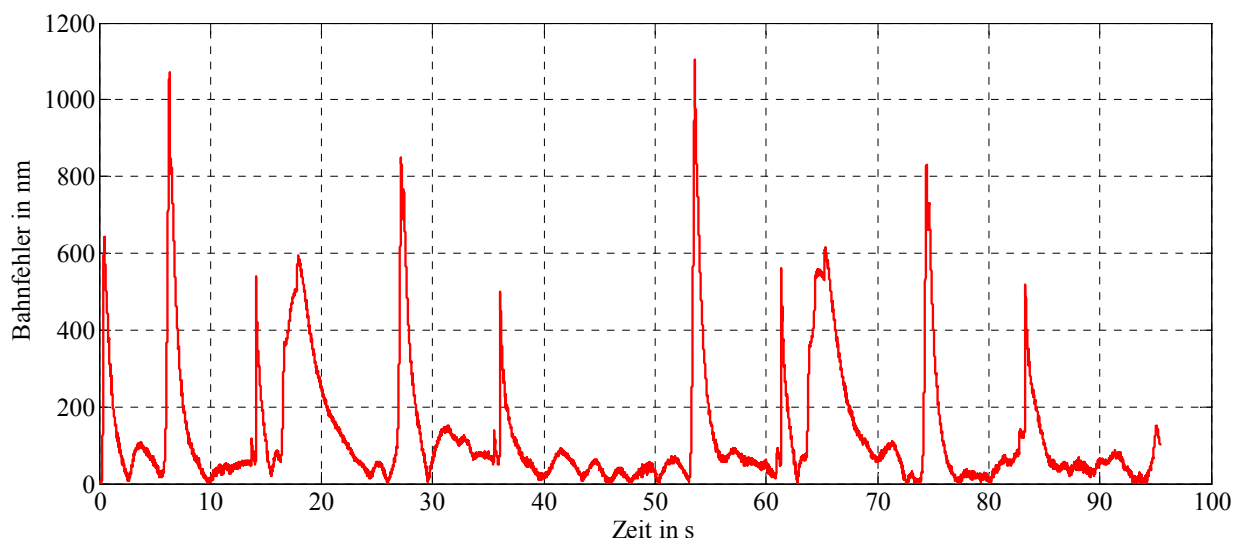


Abbildung 6.9: Bahnabweichung beim Abfahren von Figur B unter Nutzung eines dezentralen PID-Reglers

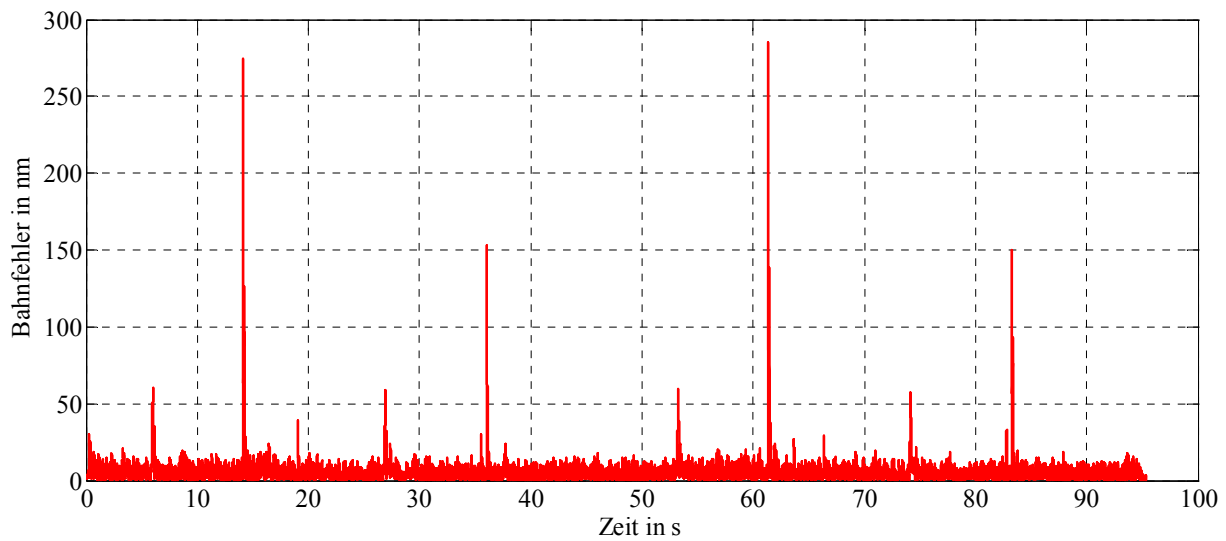


Abbildung 6.10: Bahnabweichung beim Abfahren von Figur B unter Nutzung des dezentralen modellbasierten Regelungskonzeptes

Zusammenfassend kann festgestellt werden, dass die Regelgüte auch bei komplexen zweidimensionalen Figuren durch den Einsatz von modellbasierten Einzelachsregelungen massiv verbessert werden kann. Diese Beobachtungen entsprechen den Ergebnissen aus Kapitel 5. Somit beeinflusst das vorgeschlagene dezentrale Regelungskonzept die Regelgüte nicht negativ. Des Weiteren konnte in diesem Kapitel die Funktionstüchtigkeit der entwickelten neuartigen Bahnplanungsmethode nachgewiesen werden.

7 Zusammenfassung und Ausblick

Die 3D-Nanokoordinatenmesstechnik ist geprägt von einer ständigen Vergrößerung der Messbereiche bei gleichzeitiger Beibehaltung der Positionsauflösung. Im gleichen Verhältnis wie das Messvolumen steigt, vergrößert sich natürlich der Zeitbedarf bei der Messung eines größeren Prüflings. Die resultierenden Messzeiten im zweistelligen Stundenbereich sind aus messtechnischer Sicht inakzeptabel, weshalb die Messgeschwindigkeit dieser Maschinen deutlich gesteigert werden muss. Mit dem Ziel, eine Lösungsmöglichkeit für dieses Problem zu skizzieren, wurde in der vorliegenden Arbeit erstmalig das Konzept einer dezentralen Trajektorienfolgeregelung auf NPM-Systeme übertragen.

Gegenstand der Untersuchungen war ein zweiachsiges Nanopositioniersystem mit einem Arbeitsbereich von $200 \times 200 \text{ mm}^2$ und einer Positionsauflösung von $0,1 \text{ nm}$. Für die Realisierung einer Trajektorienfolgeregelung ist es essentiell, die Systemdynamik ausgesprochen präzise zu modellieren, damit diese nach einer Modellinversion vorgesteuert werden kann. Erste Experimente anhand der Y-Achse zeigten, dass das gesamte Systemverhalten von starken Nichtlinearitäten dominiert wird. Eine tiefergehende Analyse identifizierte die Reibung in den Kugelführungen der Achsenlager als Quelle für das beobachtete Verhalten.

Somit rückte eine regelungstechnisch angepasste Reibkraftmodellierung in den Fokus des Modellbildungsprozesses und fünf aus der Literatur bekannte dynamische Reibmodelle wurden mithilfe von experimentellen Daten identifiziert. Im Zuge der Modellvalidierung stellte sich heraus, dass alle untersuchten Ansätze in der Lage sind, das grundsätzliche Reibverhalten auf Nanometerebene zufriedenstellend nachzubilden. Nichtsdestotrotz erzeugten die Reibmodelle große Modellfehler, denn auf Basis der experimentellen Daten konnte eine starke Orts- und Zeitvarianz im Systemverhalten beobachtet werden. Auf Grundlage der Validierungsergebnisse wurden das DNLRX-Reibmodell sowie ein künstliches neuronales Netz für eine spätere Verwendung in der Trajektorienfolgeregelung ausgewählt. Um die starke Orts- und Zeitvarianz des experimentellen Aufbaues zu kompensieren, wurden die ausgewählten Modelle adaptiv gestaltet. Alle linearen Modellparameter beziehungsweise Netzgewichte wurden in der späteren Anwendung in Echtzeit angepasst. So war eine exakte Abbildung des Systemverhaltens an einer beliebigen Position zu jeder Zeit mit Nanometerpräzision möglich, was bis dato als einzigartig angesehen werden kann.

Eine weitere wesentliche Komponente der Trajektorienfolgeregelung ist ein Bahnplanungsalgorithmus. Besonders bei der Bahnplanung für eine extrem präzise Positionierung in einem großen Bewegungsbereich ergeben sich außergewöhnliche Anforderungen an die Dynamik sowie die Gleichmäßigkeit (d.h. Laufruhe durch steige Differenzierbarkeit) der auszuführenden Bewegung. Aus diesem Grund wurde im Rahmen der vorliegenden Arbeit ein neuartiger Algorithmus entworfen, der speziell auf die Anforderungen von NPM-Systemen zugeschnit-

ten ist. Dieser berechnet rein analytisch für jeden Einzelachsregler des NPM-Systems stetig differenzierbare Sollwerttrajektorien für Position, Geschwindigkeit, Beschleunigung, Ruck und Ruckanstieg unter Berücksichtigung von Beschränkungen in all diesen Größen. Dabei synchronisiert der Bahnplaner die Einzelbewegungen der Achsen, sodass der Tool Center Point eine vorgegebene Kurve im dreidimensionalen Raum beschreibt. Diese Kurve kann sich aus einer beliebigen Anzahl der Streckengrundelemente Linie und Kreis zusammensetzen, und wenn es der Nutzer wünscht, werden diese Grundelemente mit erweiterten Bloss-Kurven verschliffen. So kann eine kontinuierliche Bewegung über alle Bahnsegmente hinweg realisiert werden. Darüber hinaus sollte an dieser Stelle erwähnt werden, dass im Rahmen der vorliegenden Arbeit erstmalig eine analytisch geschlossene Bahnplanungsmethodik erarbeitet wurde, die eine kontinuierliche Bahnplanung mittels beliebigen ebenen Raumkurven ermöglicht.

Neben der Vorsteuerung und dem Bahnplanungsmodul ist eine unterlagerte Regelung Bestandteil einer Trajektorienfolgeregelung. Die Aufgabe der Regelung besteht darin, sowohl exogene Störungen als auch Modellungenauigkeiten zwischen Vorsteuerung und realem System auszugleichen. Um den Einfluss des unterlagerten Reglers auf die erreichbare Regelgüte zu untersuchen, wurden zwei verschiedene Reglertypen entworfen. Zum einen war dies ein klassischer PID-Regler und zum anderen ein PI-Zustandsregler mit Führungsgrößenaufschaltung.

Zur weiteren Steigerung der Dynamik des geregelten Systems wurde die Trajektorienfolgeregelung im Verlauf der Arbeit um einen Störbeobachter erweitert. Dieser basiert auf dem Schätzalgorithmus von Kalman und berechnet den Einfluss unmodellierbarer Systemgrößen wie beispielsweise Temperaturschwankungen.

Nach dem Entwurf der verschiedenen Regelungskomponenten wurde im experimentellen Teil der Arbeit die Leistungsfähigkeit dieser Module untersucht. Die Ermittlung der optimalen Reglerkonfiguration erfolgte zu Beginn der Arbeiten anhand der Y-Achse. Hierzu wurde die Regelgüte verschiedener Ansätze gegenübergestellt und bewertet. Während dieser Analyse zeigte sich, dass eine Kombination aus dem adaptiven DNLRX-Reibmodell, Zustandsregler und dem entworfenen Störbeobachter einen optimalen Kompromiss aus Regelgüte und Rechenzeit darstellt. Im Vergleich zu klassisch motivierten Regelungen konnte mit diesem Ansatz die dynamische Regelabweichung im gesamten Arbeitsbereich der Maschine signifikant verbessert werden.

Nach den positiven Ergebnissen wurde das Konzept der Trajektorienfolgeregelung auf die zweite Bewegungsachse des xy-Demonstrators übertragen und ein dezentraler Regelungsansatz umgesetzt. Hierzu war es notwendig, eine Trajektorienfolgeregelung auch für die X-Achse zu entwerfen und beide Einzelachsregelungen in Kombination mit dem entwickelten Bahnplanungsmodul in Betrieb zu nehmen. Im Anschluss daran wurden auch hier anhand mehrerer Experimente sehr gute Resultate erzielt. Die Funktionstüchtigkeit sowohl der modellbasierten Achsregelungen als auch des entwickelten Bahnplanungsalgorithmus konnte damit nachgewiesen werden, und es zeigte sich sehr deutlich, dass der gewählte Regelungsansatz die Dynamik von NPM-Systemen signifikant steigern kann.

Im Rahmen dieser Arbeit konnte nachgewiesen werden, dass ein modellbasierter Regelungsansatz auch bei NPM-Systemen sehr großes Potential zur Minimierung des dynamischen Positionsfehlers bietet. Die Kernkomponente eines derartigen Ansatzes ist ein möglichst exaktes Modell der Maschinendynamik und daher sollte der Schwerpunkt zukünftiger Arbeiten auf einer weiteren Verbesserung dieser Modelle liegen. Darüber hinaus kann der vorgestellte Bahnplanungsalgorithmus weiterentwickelt werden. Die rotatorischen Freiheitsgrade des Systems wurden bisher noch nicht berücksichtigt und so liegt es nahe, den Algorithmus zu einem sechs-Achsen-Bahnplaner zu erweitern. Auch besteht weiterhin erhebliches Forschungspotential bei der Berechnung der Bahnbeschränkungen für gekrümmte Raumkurven. Deren Bestimmung erfolgt zurzeit mittels numerischer Methoden, was aus mathematischer Sicht eher suboptimal ist.

Sind alle vorgeschlagenen algorithmischen Möglichkeiten ausgeschöpft, kann nur noch der experimentelle Aufbau weiterentwickelt werden. Während der Untersuchungen zeigte sich, dass die Verwendung zusätzlicher Messgrößen die Qualität der Regelung stark verbessert. Besonders eine messtechnische Erfassung der Geschwindigkeit kann die Regelgüte weiter verbessern, denn erst dann wird das gesamte Potential der Zustandsregelung ausgeschöpft.

Natürlich wird eine weitere Optimierung der Antriebstechnik und des mechanischen Aufbaues die erreichbare Regelgüte weiter verbessern. Dies liegt jedoch nicht im Fokus der vorliegenden Arbeit und soll daher nicht weiter ausgeführt werden.

Anhang

Anhang A

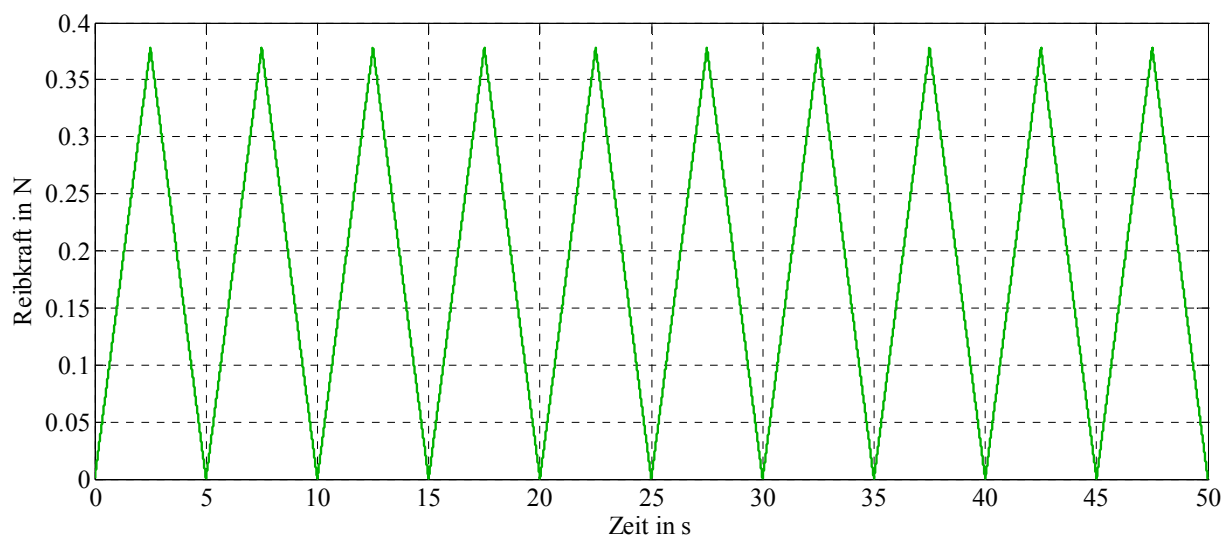


Abbildung A 1: Zeitlicher Verlauf der Antriebskraft bei einer Bewegung im „Pre-Rolling“-Bereich

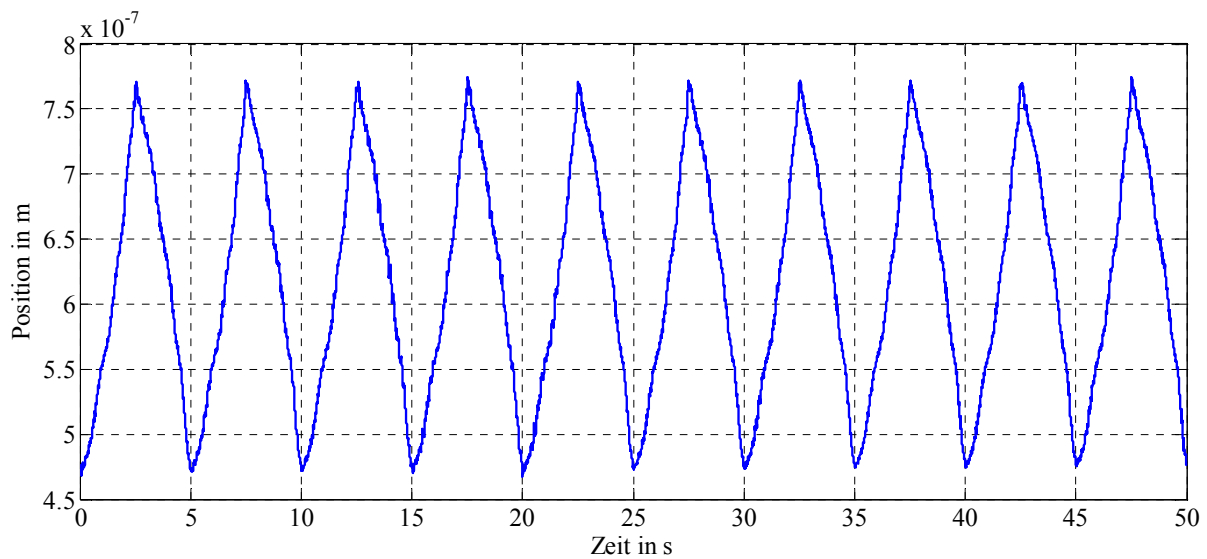


Abbildung A 2: Zeitlicher Verlauf der Position bei einer Bewegung im „Pre-Rolling“-Bereich

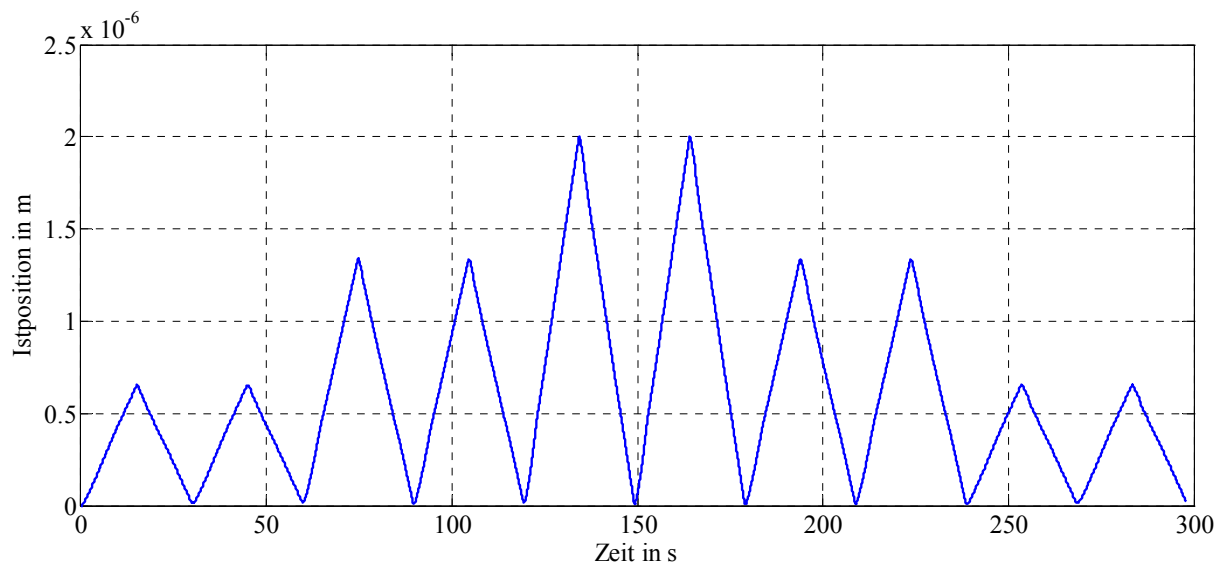


Abbildung A 3: Zeitlicher Verlauf der Position bei Bewegungen im „Pre-Rolling“-Bereich mit unterschiedlichen Amplituden

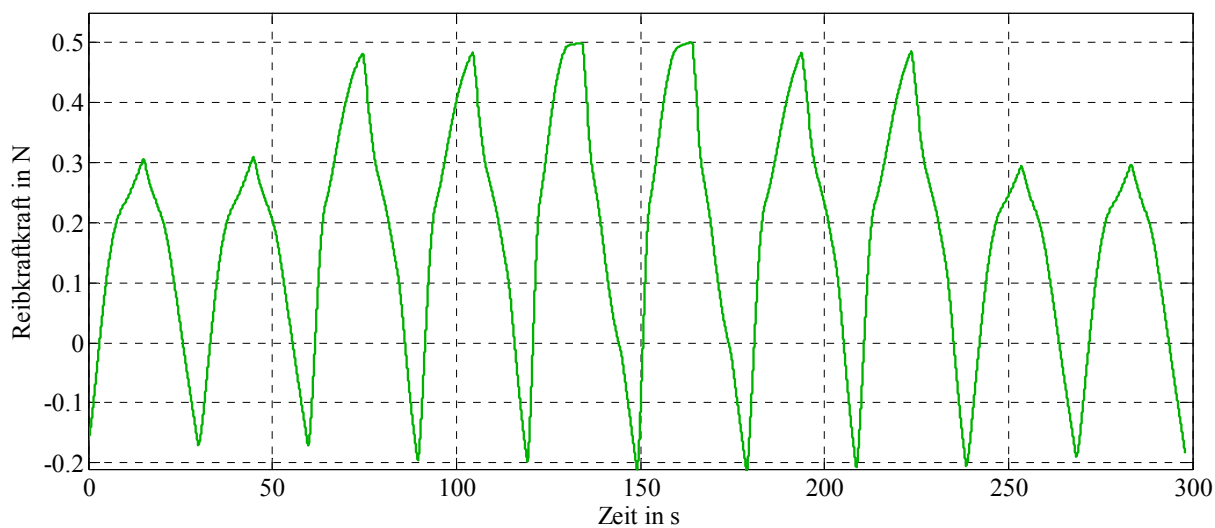


Abbildung A 4: Zeitlicher Verlauf der Antriebskraft bei Bewegungen im „Pre-Rolling“-Bereich mit unterschiedlichen Amplituden

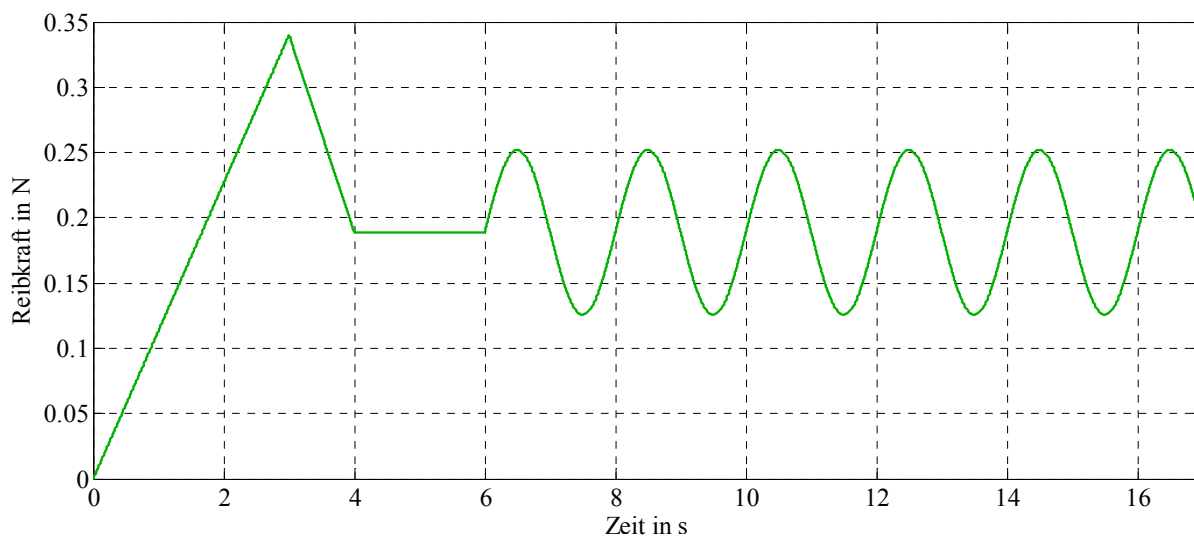


Abbildung A 5: Zeitlicher Verlauf der Antriebskraft bei einer Bewegung mit reversiblen und irreversiblen Anteilen

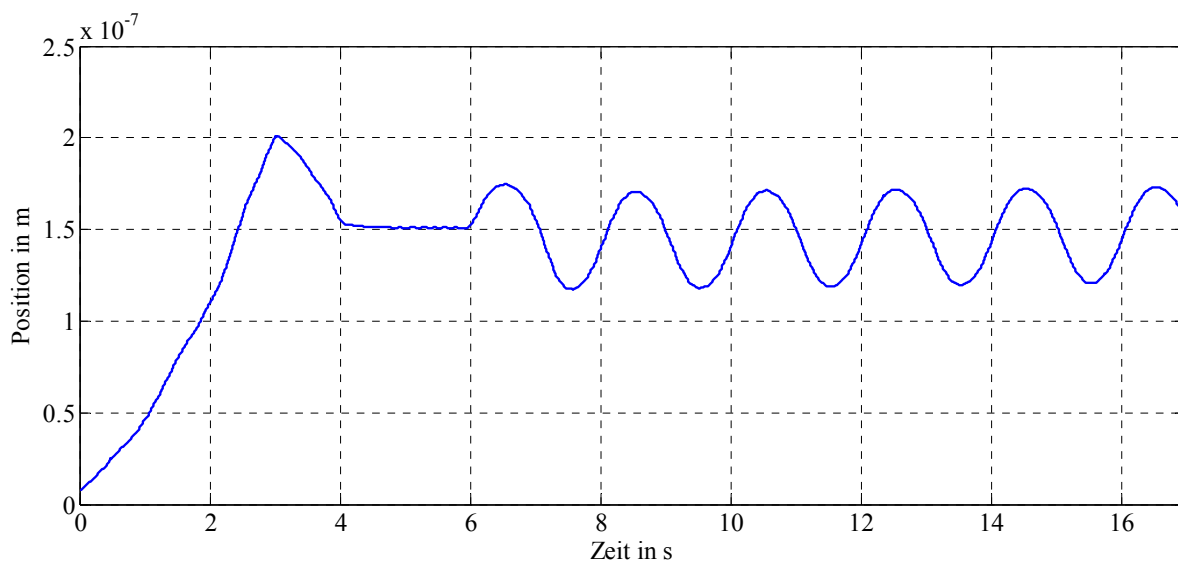


Abbildung A 6: Zeitlicher Verlauf der Position bei einer Bewegung mit reversiblen und irreversiblen Anteilen

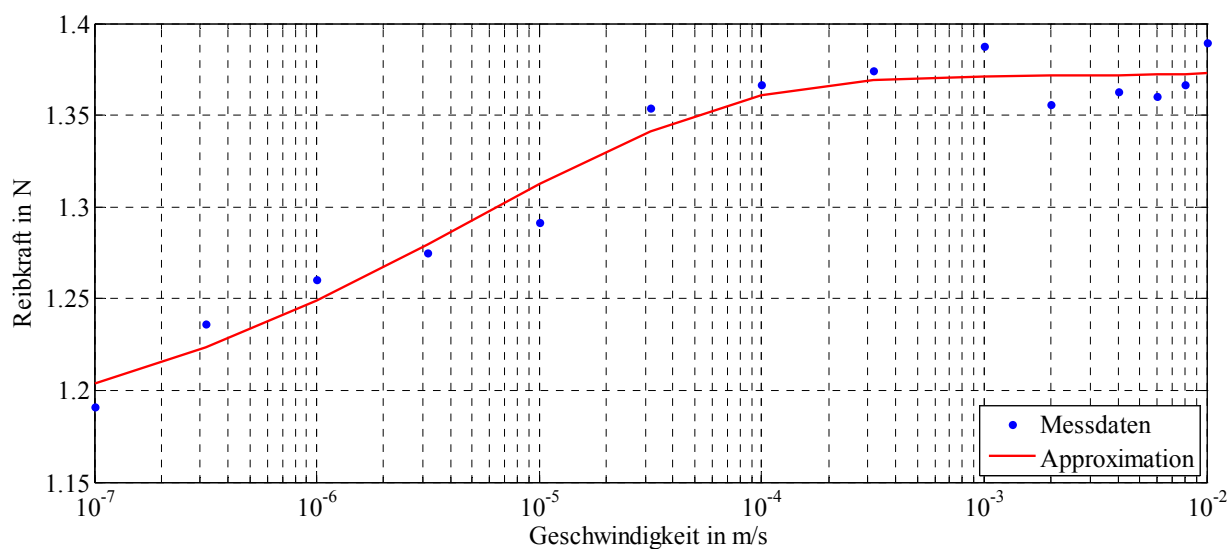


Abbildung A 7: Logarithmische Darstellung der statischen Reibkennlinie für positive Geschwindigkeiten

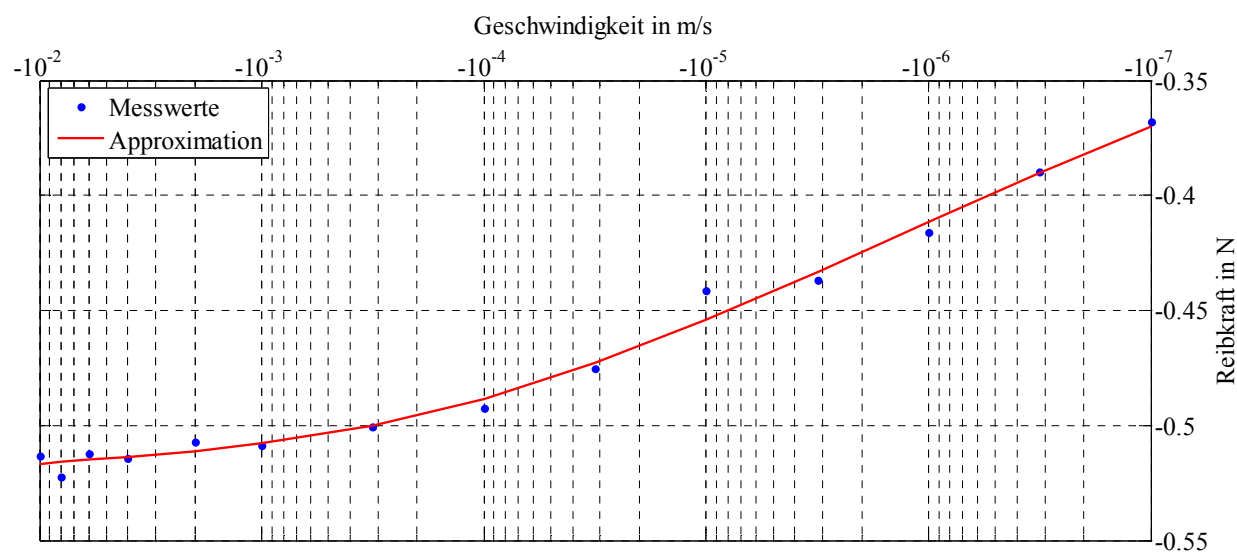


Abbildung A 8: Logarithmische Darstellung der statischen Reibkennlinie für negative Geschwindigkeiten

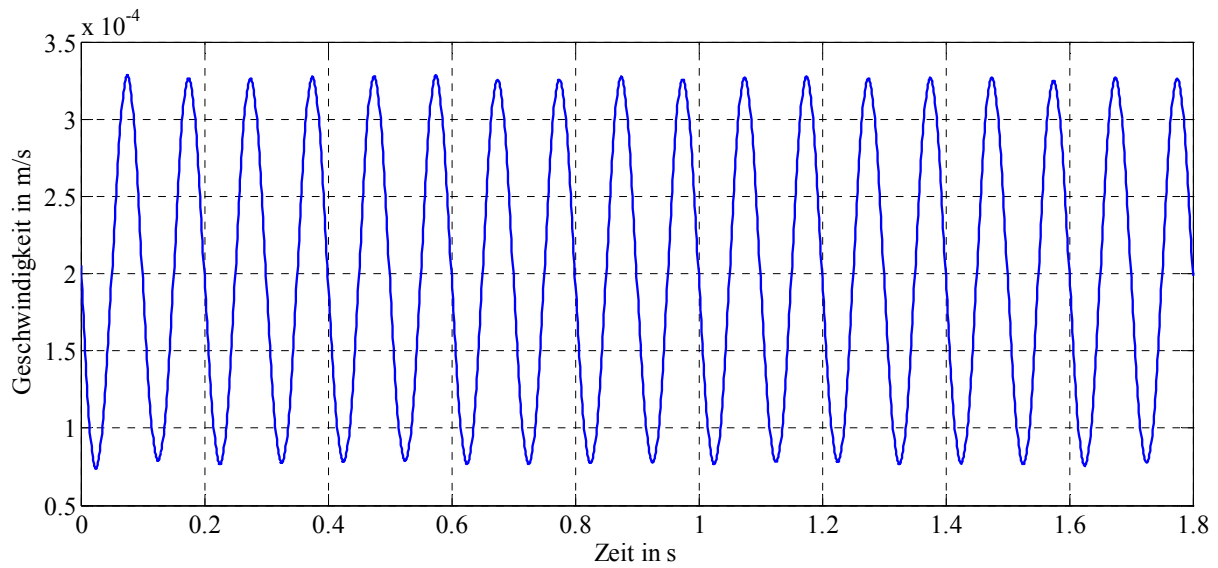


Abbildung A 9: Zeitlicher Verlauf der Geschwindigkeit zum Nachweis des „frictional lag“-Effektes

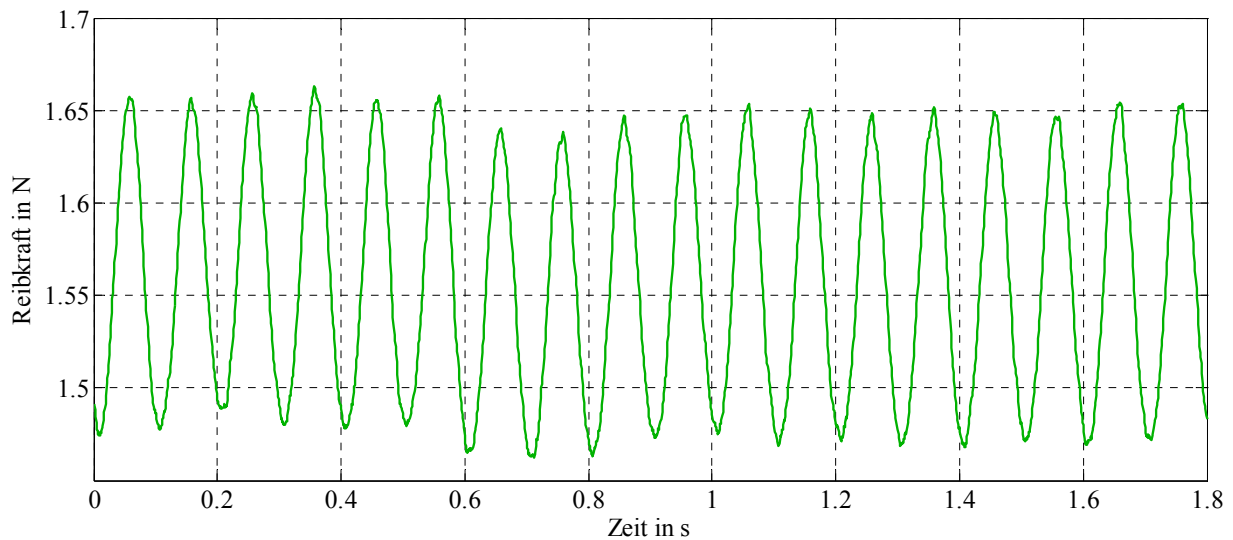


Abbildung A 10: Zeitlicher Verlauf der Reibkraft zum Nachweis des „frictional lag“-Effektes

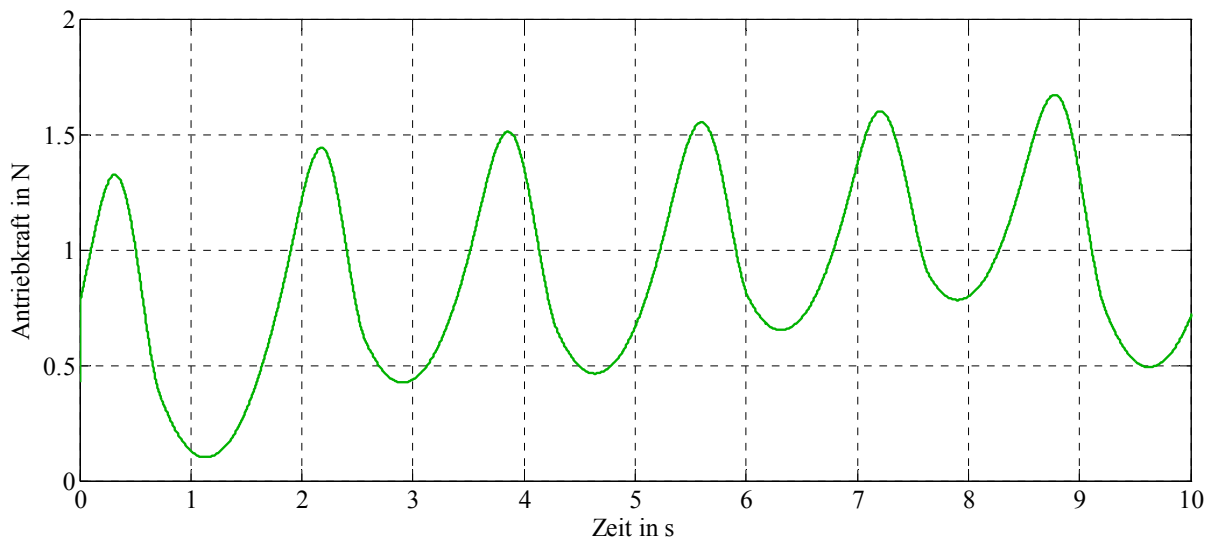


Abbildung A 11: Zeitlicher Verlauf der Antriebskraft bei Auftreten des Stick-Slip-Phänomens

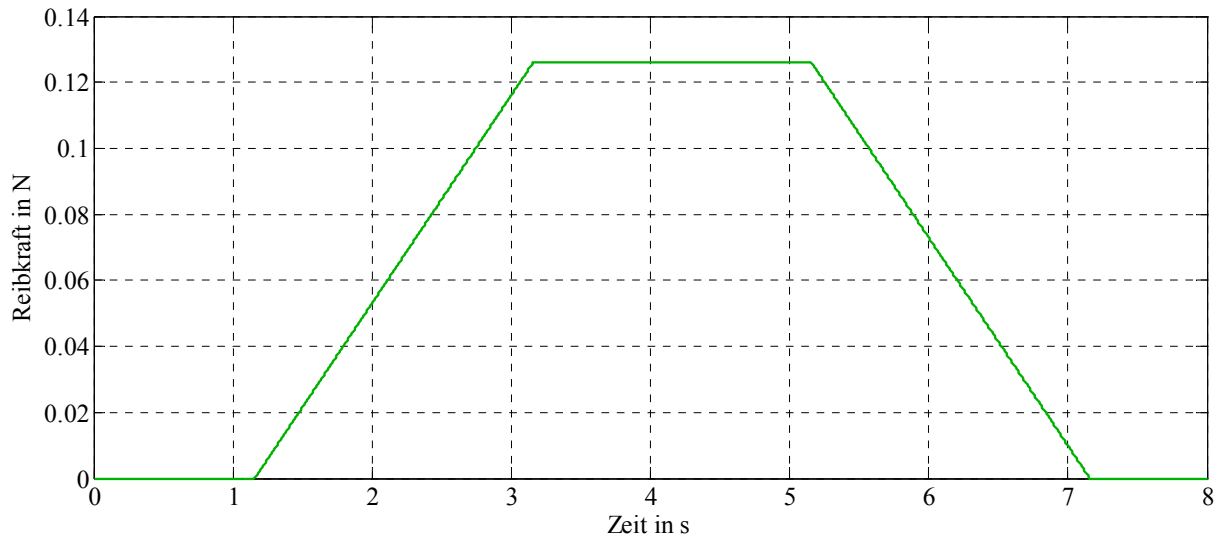


Abbildung A 12: Zeitlicher Verlauf der Antriebskraft bei einer Bewegung im „Pre-Rolling“-Bereich zur Bestimmung von σ_1 , σ_2 und z_{ba}

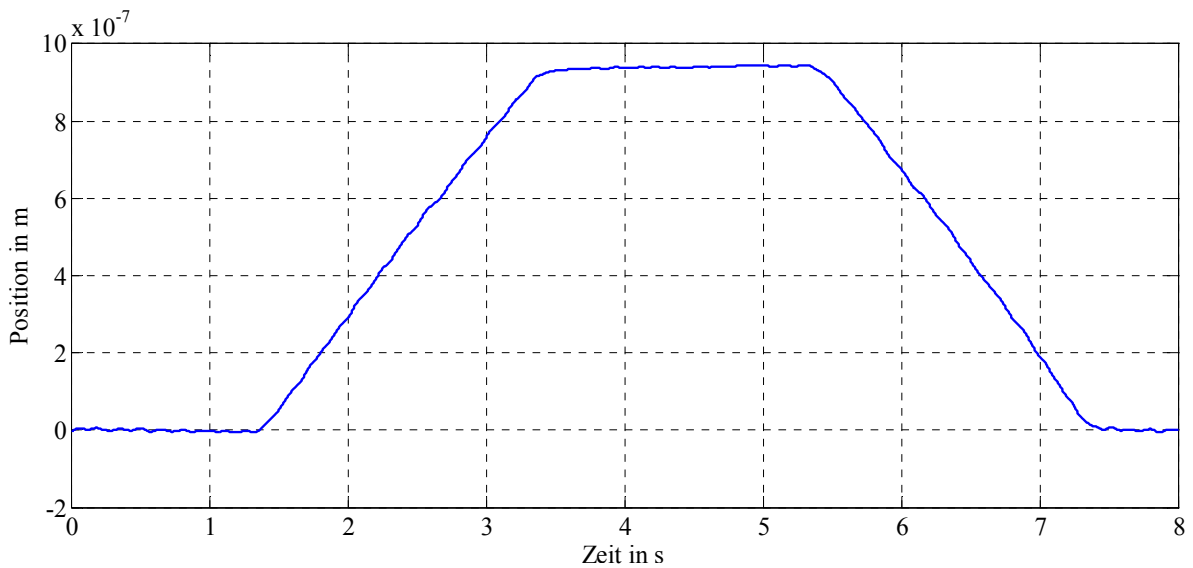


Abbildung A 13: Zeitlicher Verlauf der Position bei einer Bewegung im „Pre-Rolling“-Bereich zur Bestimmung von σ_1 , σ_2 und z_{ba}

Anhang B

Zur Lösung von Ungleichung (4.104) kann das Optimierungsproblem folgendermaßen formuliert werden:

OP 1: $\min_{s, v, a, j} |v|$, sodass

$$\sqrt{(\kappa'v^3 + 3\kappa va)^2 + (j - \kappa^2 v^3)^2} - \hat{j}_{max} = 0$$

$$-s \leq 0$$

$$s - s_{ziel} \leq 0$$

$$a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$-j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

OP 2: $\min_{s, v, a, j} |a|$, sodass

$$\sqrt{(\kappa'v^3 + 3\kappa va)^2 + (j - \kappa^2 v^3)^2} - \hat{j}_{max} = 0$$

$$-s \leq 0$$

$$s - s_{ziel} \leq 0$$

$$v - \frac{\hat{v}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

$$-v - \frac{\hat{v}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

$$j - \frac{\hat{j}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

$$-j - \frac{\hat{j}_{max}}{\hat{a}_{max}} \cdot |a| \leq 0$$

OP 3: $\min_{s, v, a, j} |j|$, sodass

$$\sqrt{(\kappa'v^3 + 3\kappa va)^2 + (j - \kappa^2 v^3)^2} - \hat{j}_{max} = 0$$

$$-s \leq 0$$

$$s - s_{ziel} \leq 0$$

$$v - \frac{\hat{v}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$-v - \frac{\hat{v}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$a - \frac{\hat{a}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{j}_{max}} \cdot |j| \leq 0$$

In gleicher Art und Weise können auch die Optimierungsprobleme zur Lösung von Ungleichung (4.105) angepasst werden:

OP 1: $\min_{s, v, a, j, d} |v|$, sodass

$$\sqrt{(3\kappa a^2 + 4\kappa vj - \kappa^3 v^4 + 6\kappa'v^2 a + \kappa''v^4)^2 + (d - 6\kappa^2 v^2 a - 3 \cdot \kappa \kappa' v^4)^2} - \hat{d}_{max} = 0$$

$$-s \leq 0$$

$$s - s_{ziel} \leq 0$$

$$a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$-j - \frac{\hat{j}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$d - \frac{\hat{d}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

$$-d - \frac{\hat{d}_{max}}{\hat{v}_{max}} \cdot |v| \leq 0$$

OP 2: $\min_{s, v, a, j, d} |a|$, sodass

$$\begin{aligned} & \sqrt{\left(3\kappa a^2 + 4\kappa v j - \kappa^3 v^4 + 6\kappa' v^2 a + \kappa'' v^4\right)^2 + \left(d - 6\kappa^2 v^2 a - 3 \cdot \kappa \kappa' v^4\right)^2} - \hat{d}_{\max} = 0 \\ & -s \leq 0 \\ & s - s_{\text{ziel}} \leq 0 \\ & v - \frac{\hat{v}_{\max}}{\hat{a}_{\max}} \cdot |a| \leq 0 \\ & -v - \frac{\hat{v}_{\max}}{\hat{a}_{\max}} \cdot |a| \leq 0 \\ & j - \frac{\hat{j}_{\max}}{\hat{a}_{\max}} \cdot |a| \leq 0 \\ & -j - \frac{\hat{j}_{\max}}{\hat{a}_{\max}} \cdot |a| \leq 0 \\ & d - \frac{\hat{d}_{\max}}{\hat{a}_{\max}} \cdot |a| \leq 0 \\ & -d - \frac{\hat{d}_{\max}}{\hat{a}_{\max}} \cdot |a| \leq 0 \end{aligned}$$

OP 3: $\min_{s, v, a, j, d} |j|$, sodass

$$\begin{aligned} & \sqrt{\left(3\kappa a^2 + 4\kappa v j - \kappa^3 v^4 + 6\kappa' v^2 a + \kappa'' v^4\right)^2 + \left(d - 6\kappa^2 v^2 a - 3 \cdot \kappa \kappa' v^4\right)^2} - \hat{d}_{\max} = 0 \\ & -s \leq 0 \\ & s - s_{\text{ziel}} \leq 0 \\ & v - \frac{\hat{v}_{\max}}{\hat{j}_{\max}} \cdot |j| \leq 0 \\ & -v - \frac{\hat{v}_{\max}}{\hat{j}_{\max}} \cdot |j| \leq 0 \\ & a - \frac{\hat{a}_{\max}}{\hat{j}_{\max}} \cdot |j| \leq 0 \\ & -a - \frac{\hat{a}_{\max}}{\hat{j}_{\max}} \cdot |j| \leq 0 \\ & d - \frac{\hat{d}_{\max}}{\hat{j}_{\max}} \cdot |j| \leq 0 \\ & -d - \frac{\hat{d}_{\max}}{\hat{j}_{\max}} \cdot |j| \leq 0 \end{aligned}$$

OP 4: $\min_{s, v, a, j, d} |d|$, sodass

$$\sqrt{\left(3\kappa a^2 + 4\kappa v j - \kappa^3 v^4 + 6\kappa' v^2 a + \kappa'' v^4\right)^2 + \left(d - 6\kappa^2 v^2 a - 3 \cdot \kappa \kappa' v^4\right)^2} - \hat{d}_{max} = 0$$

$$-s \leq 0$$

$$s - s_{ziel} \leq 0$$

$$v - \frac{\hat{v}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$-v - \frac{\hat{v}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$a - \frac{\hat{a}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$-a - \frac{\hat{a}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$j - \frac{\hat{j}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

$$-j - \frac{\hat{j}_{max}}{\hat{d}_{max}} \cdot |d| \leq 0$$

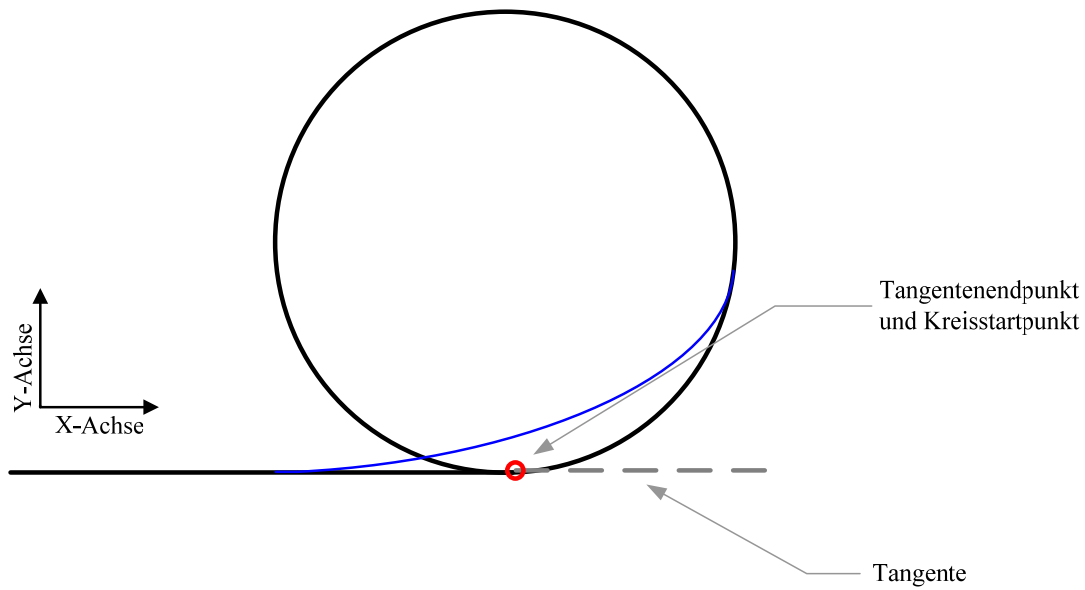


Abbildung A 14: Beispiel für das Verschleifen von Tangente (ohne tangentiale Abrückung) und Kreis

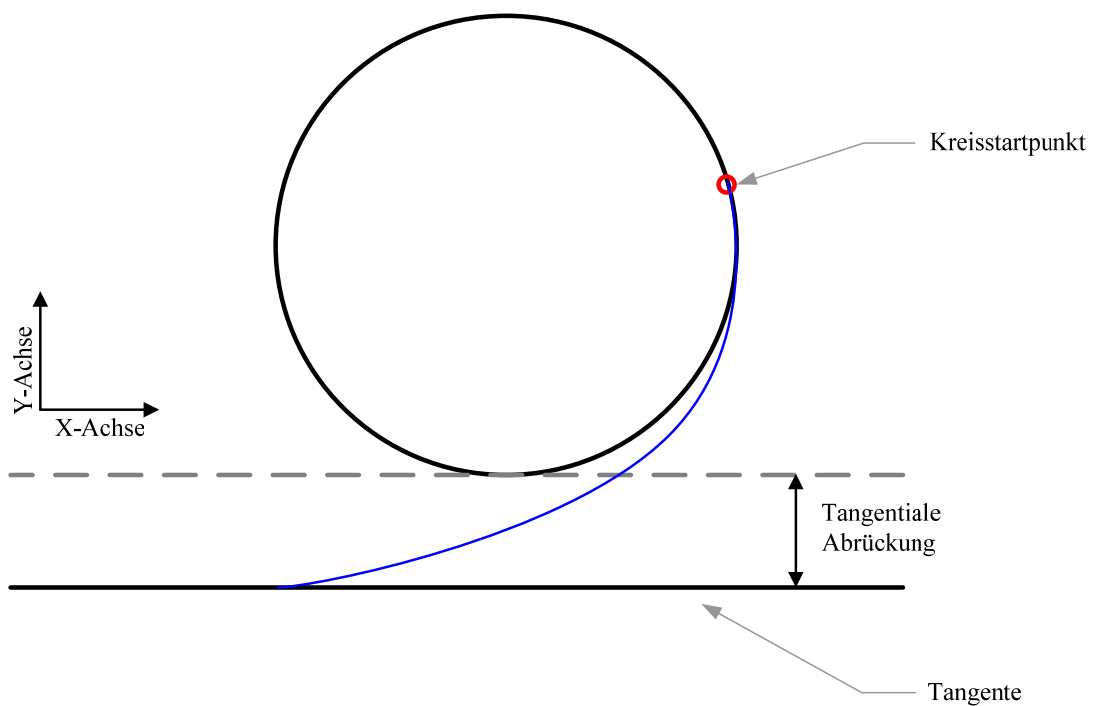


Abbildung A 15: Beispiel für das Verschleifen von Tangente (mit tangentialer Abrückung) und Kreis

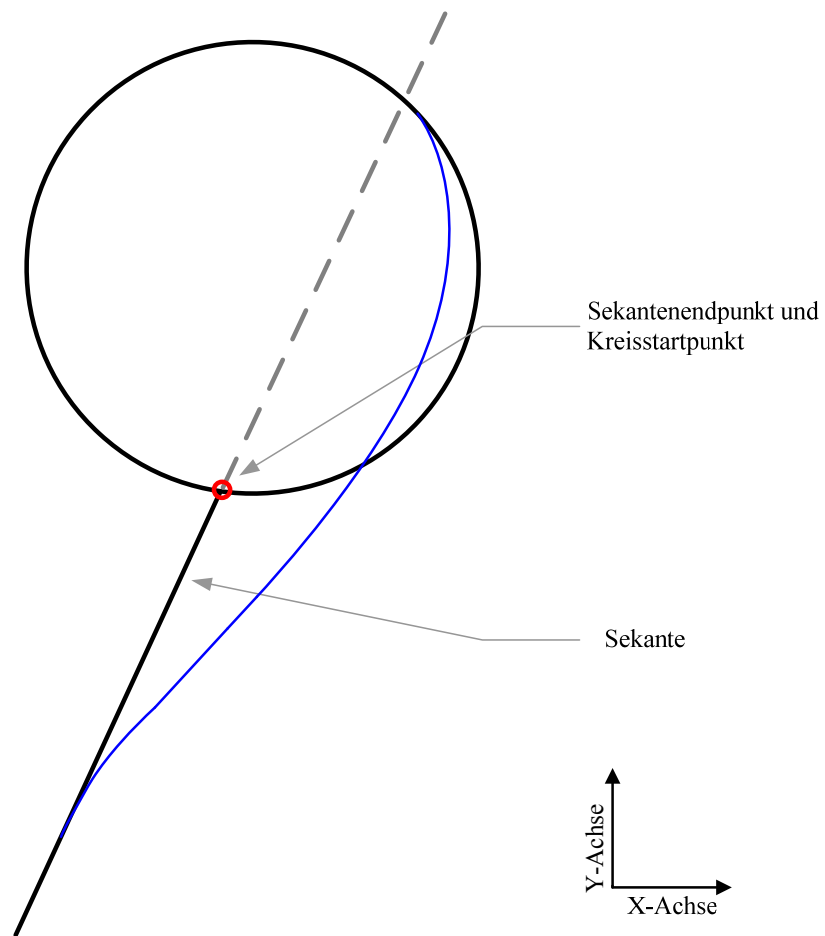


Abbildung A 16: Beispiel für das Verschleifen von Sekante und Kreis

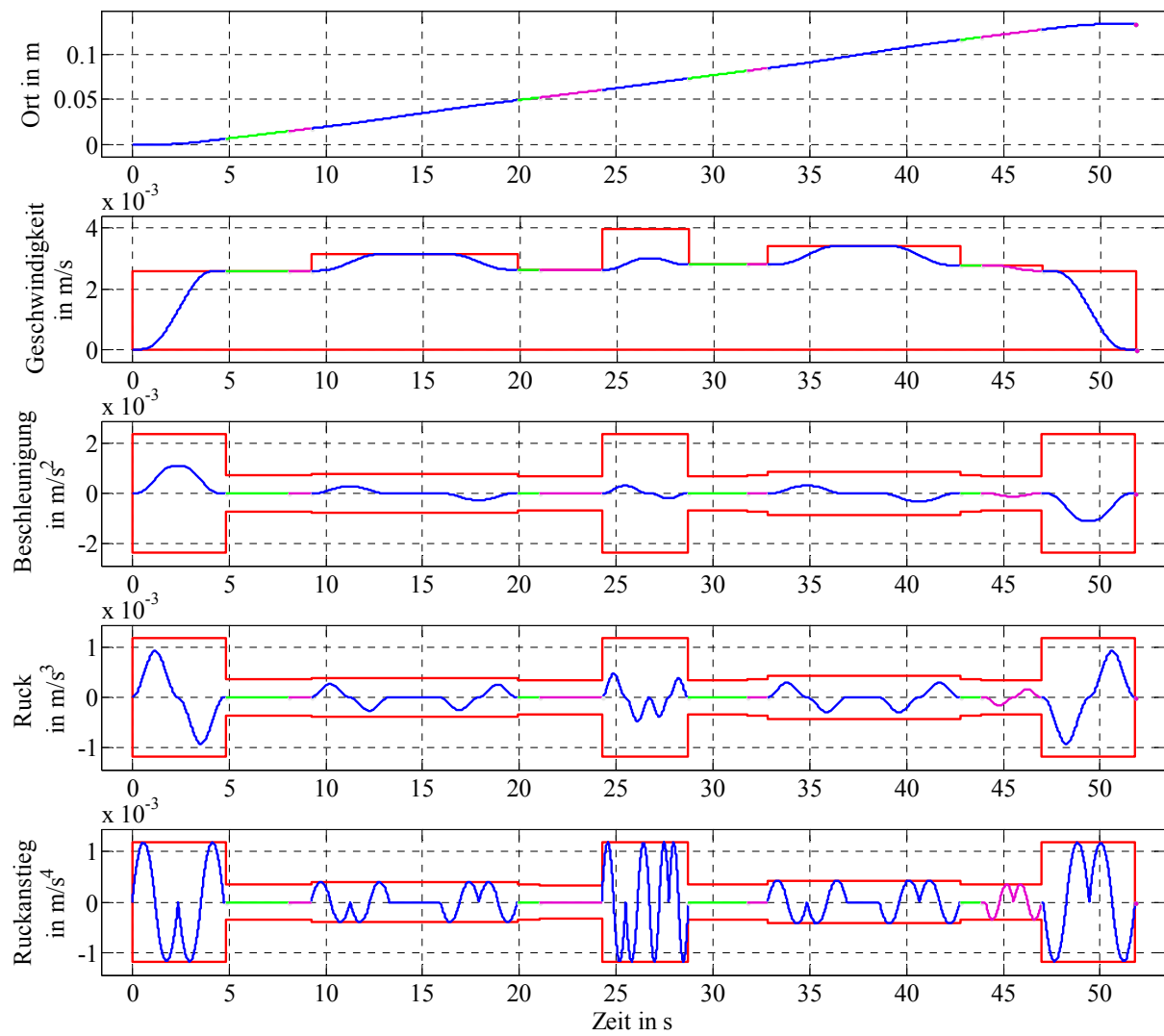
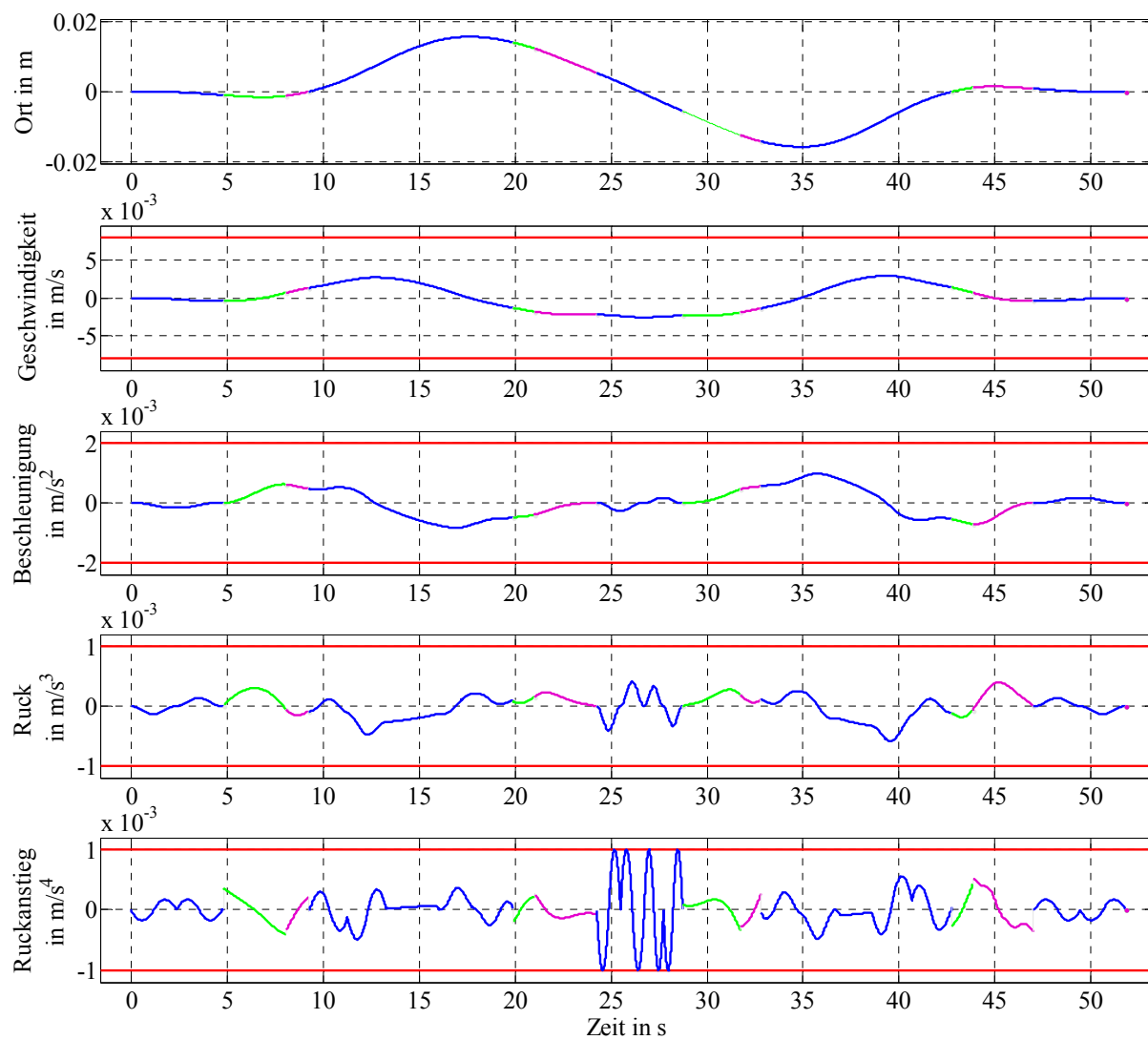


Abbildung A 17: Kinematik auf Bahnebene

**Abbildung A 18: Kinematik auf der ersten Koordinatenachse**

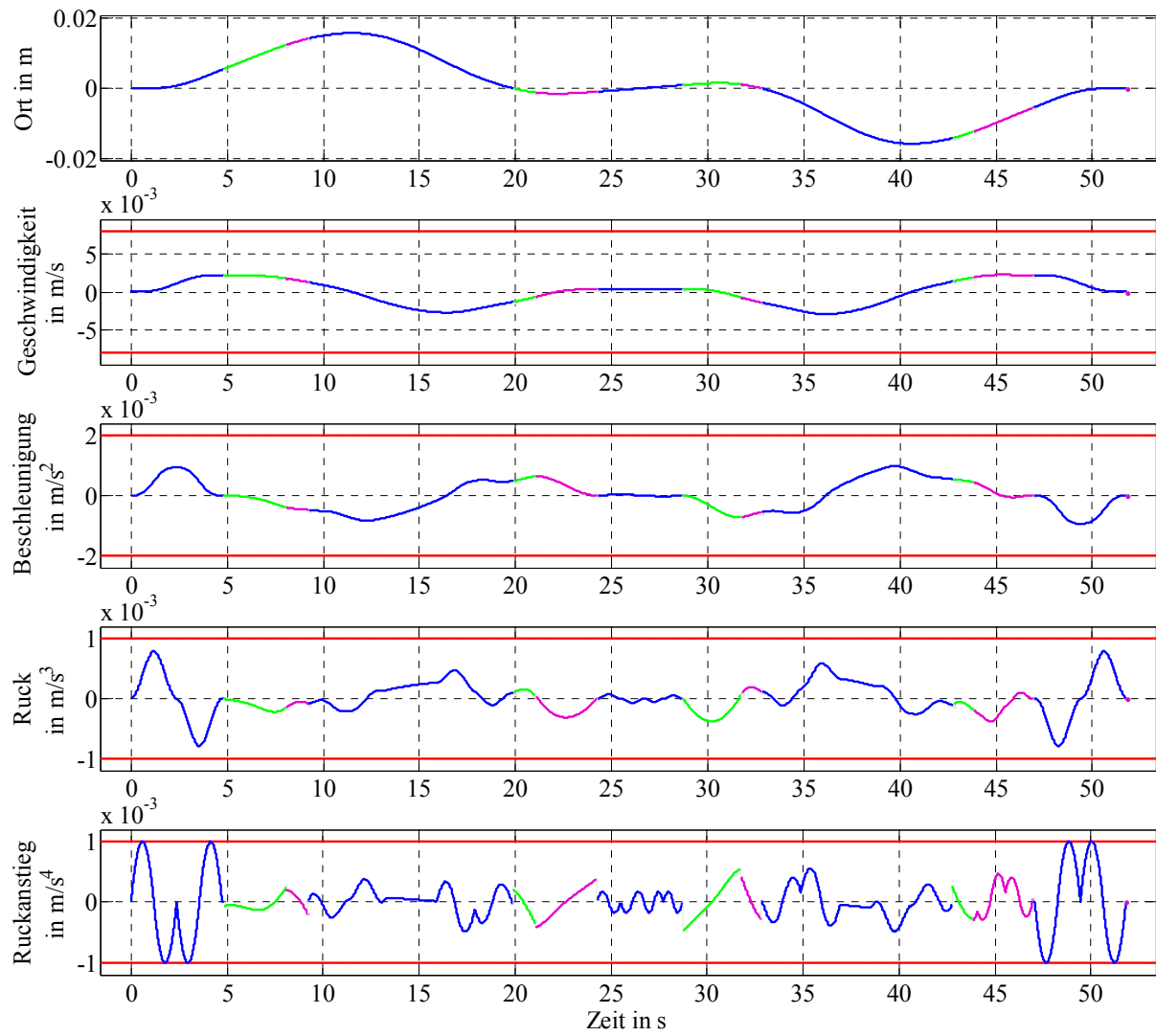


Abbildung A 19: Kinematik auf der zweiten Koordinatenachse

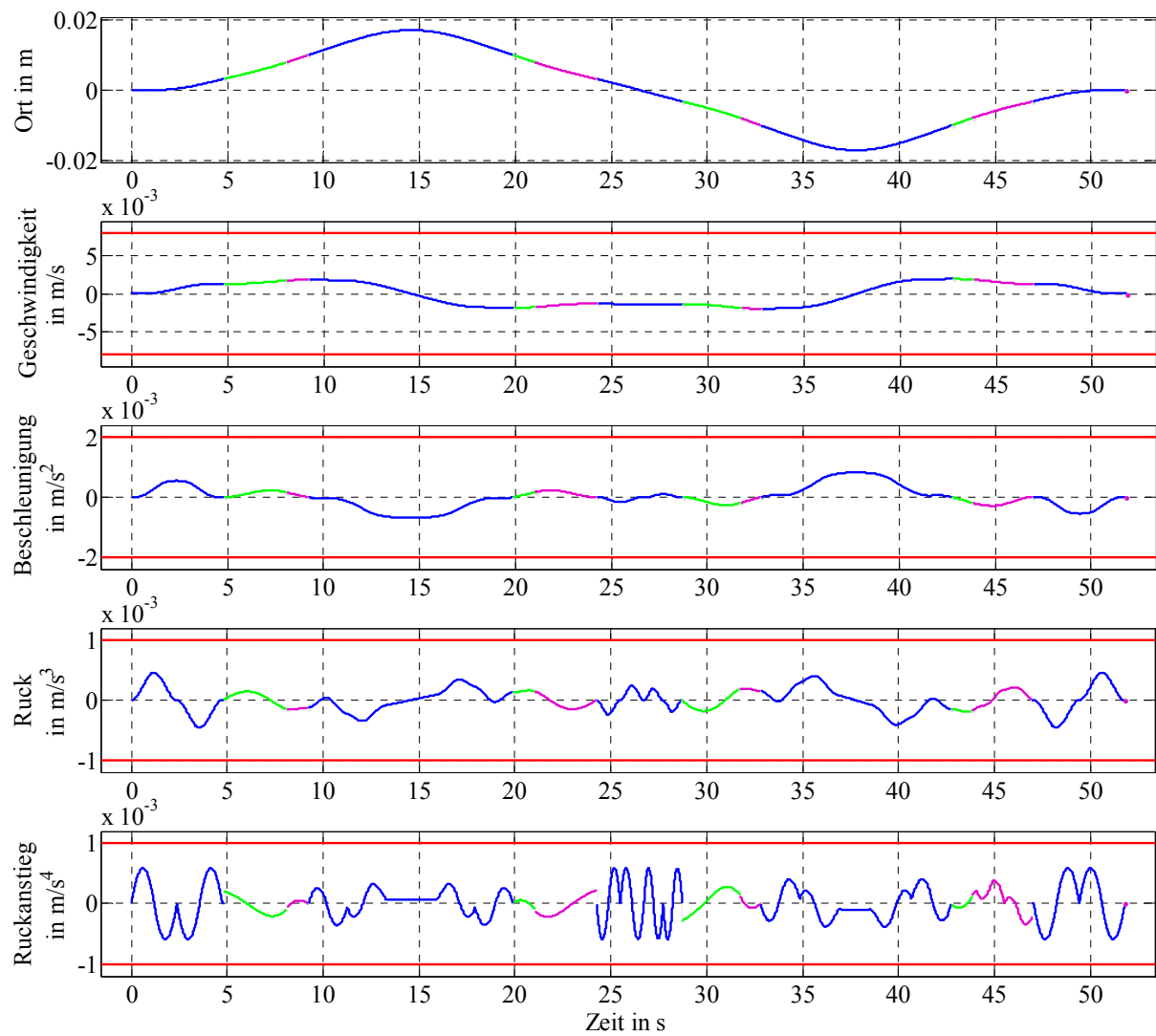


Abbildung A 20: Kinematik auf der dritten Koordinatenachse

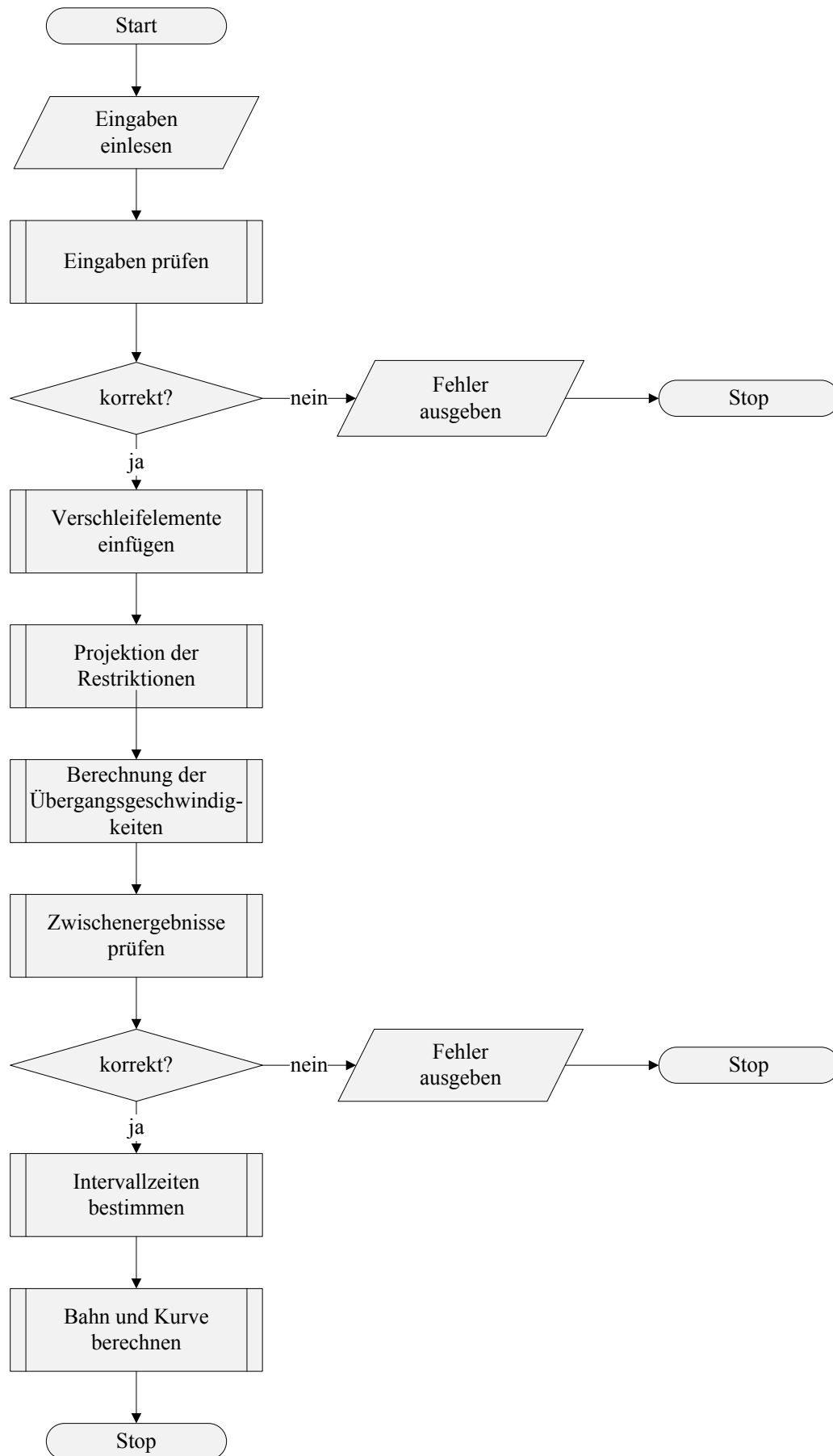


Abbildung A 21: Gesamtstruktur der Bahnplanung

Anhang C

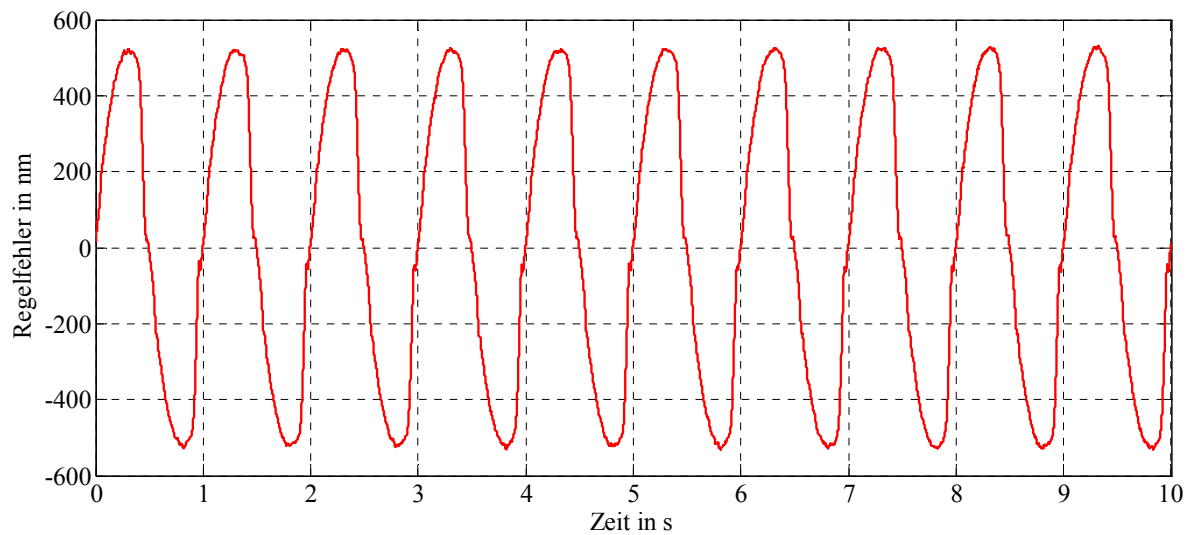


Abbildung A 22: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines klassischen PID-Reglers

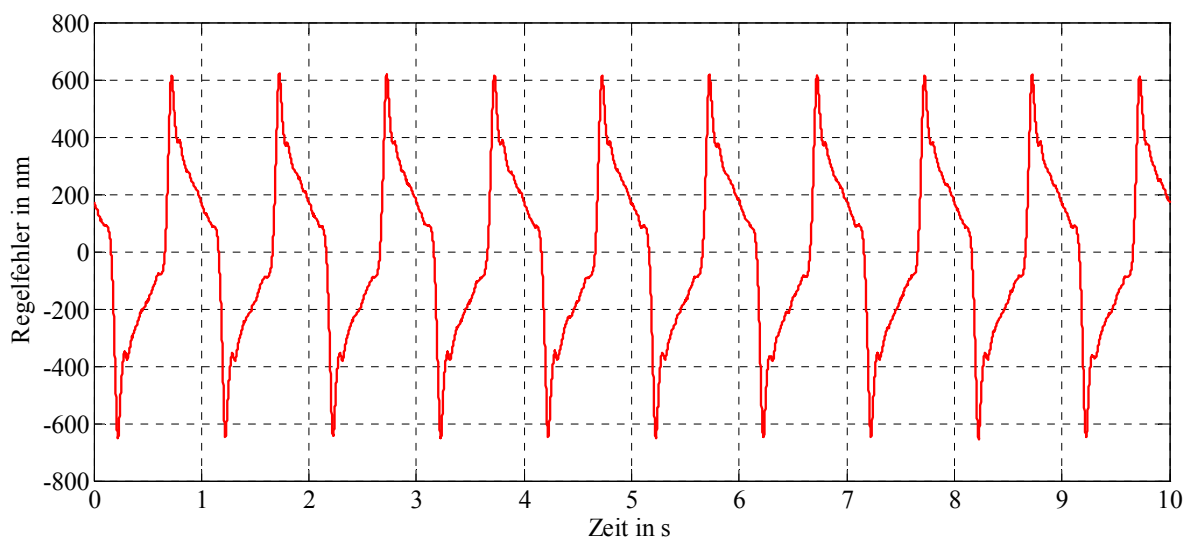


Abbildung A 23: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines PI-Zustandsreglers

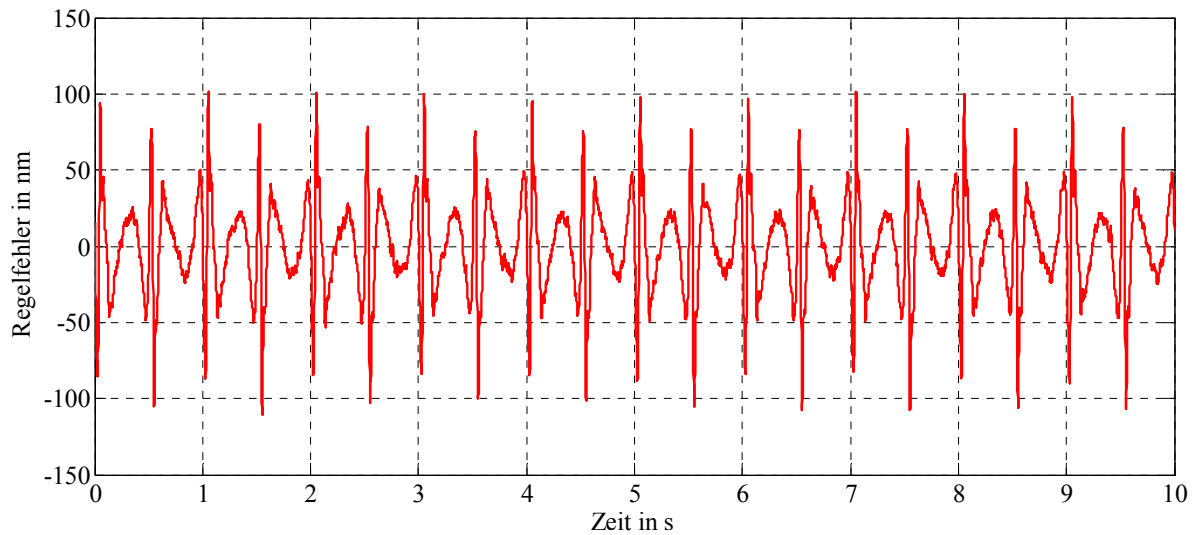


Abbildung A 24: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung

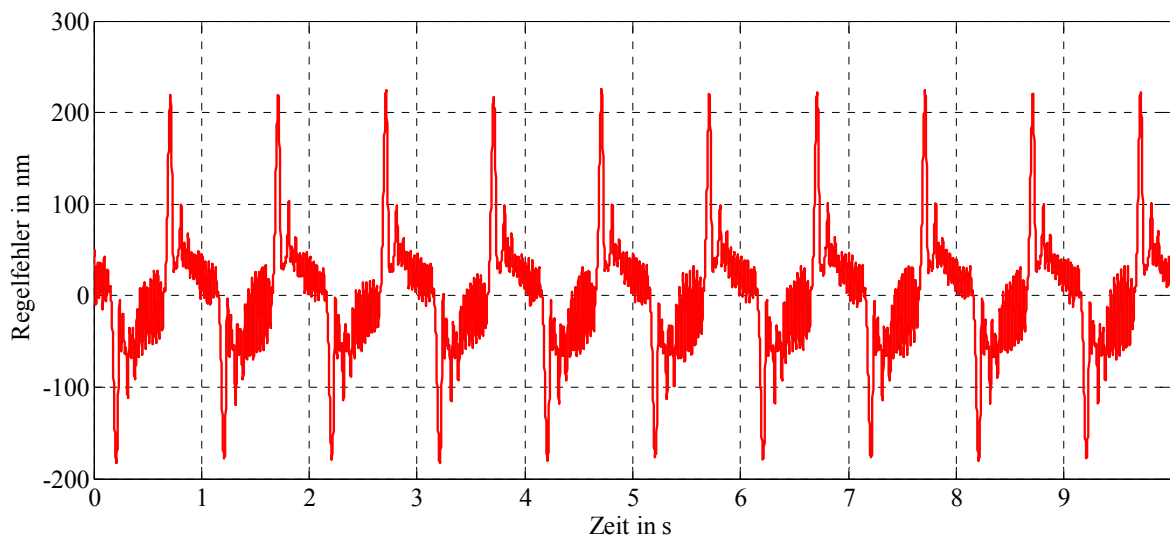


Abbildung A 25: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem KNN als Vorsteuerung

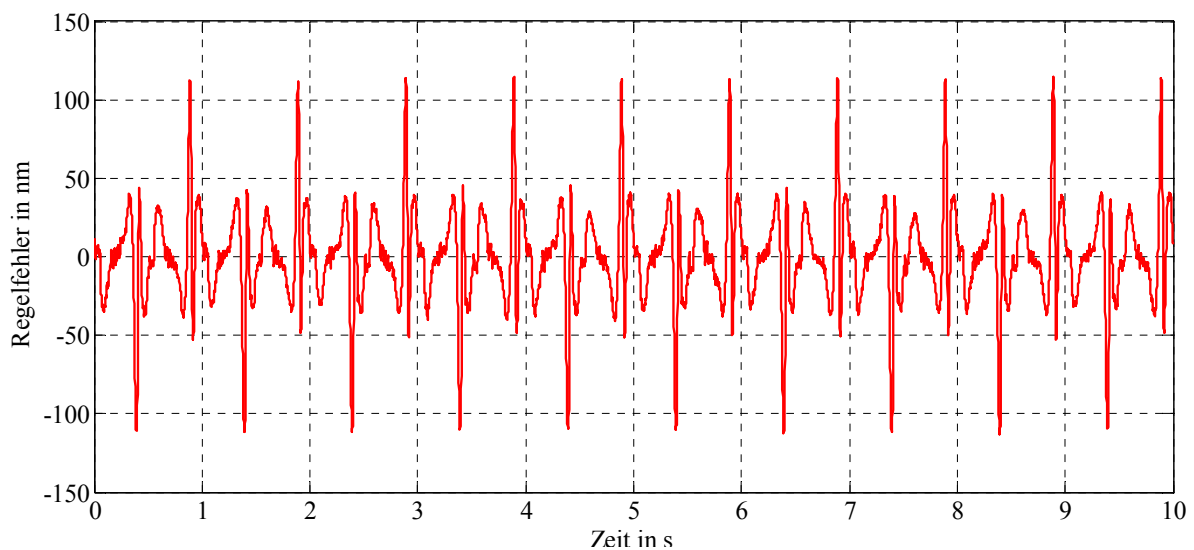


Abbildung A 26: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung

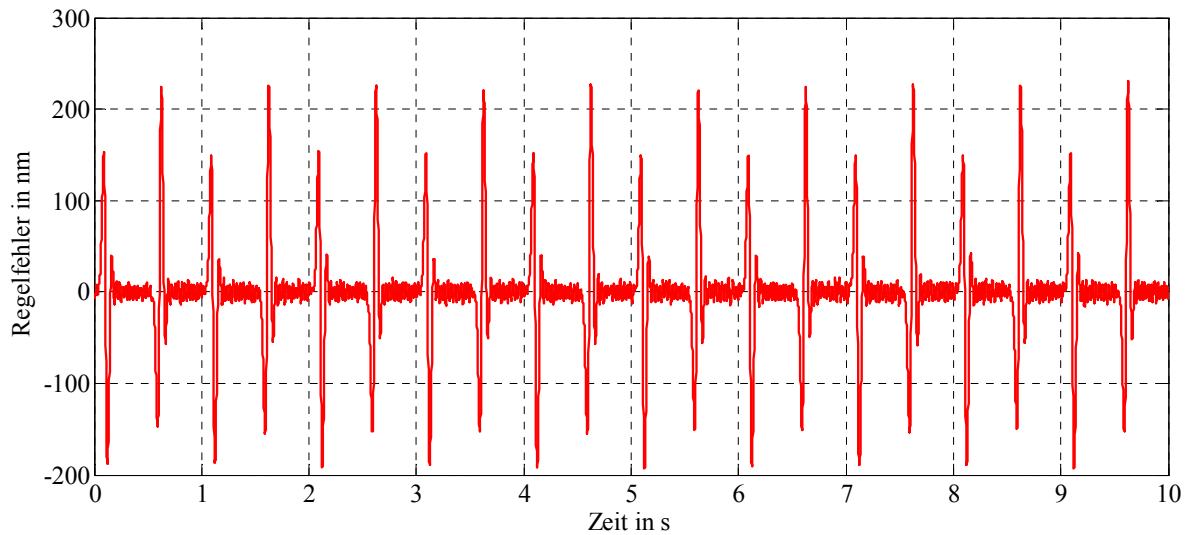


Abbildung A 27: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem KNN als Vorsteuerung

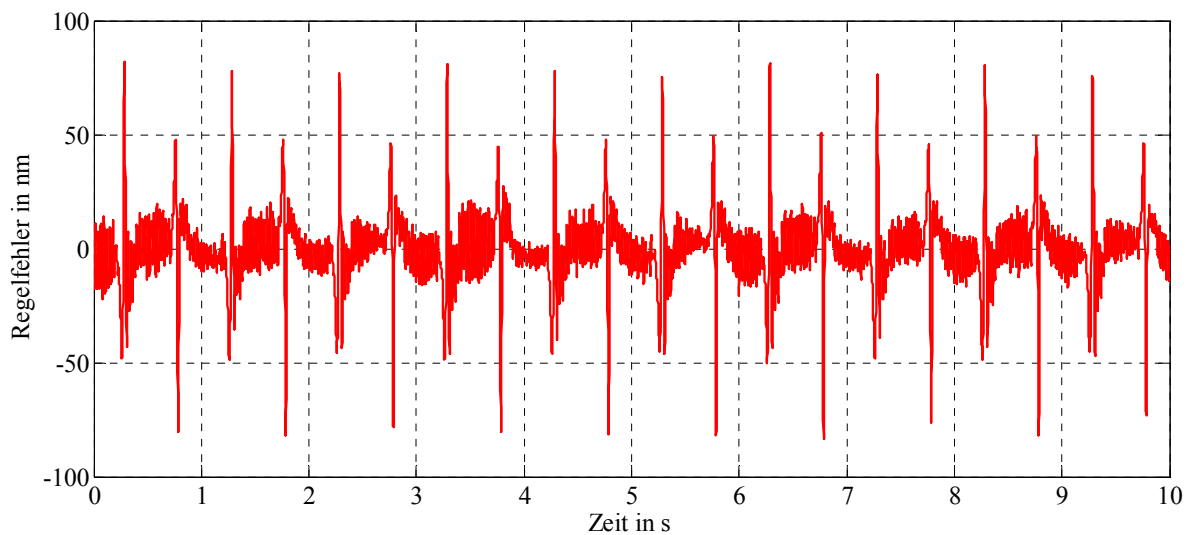


Abbildung A 28: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung und einem Störbeobachter

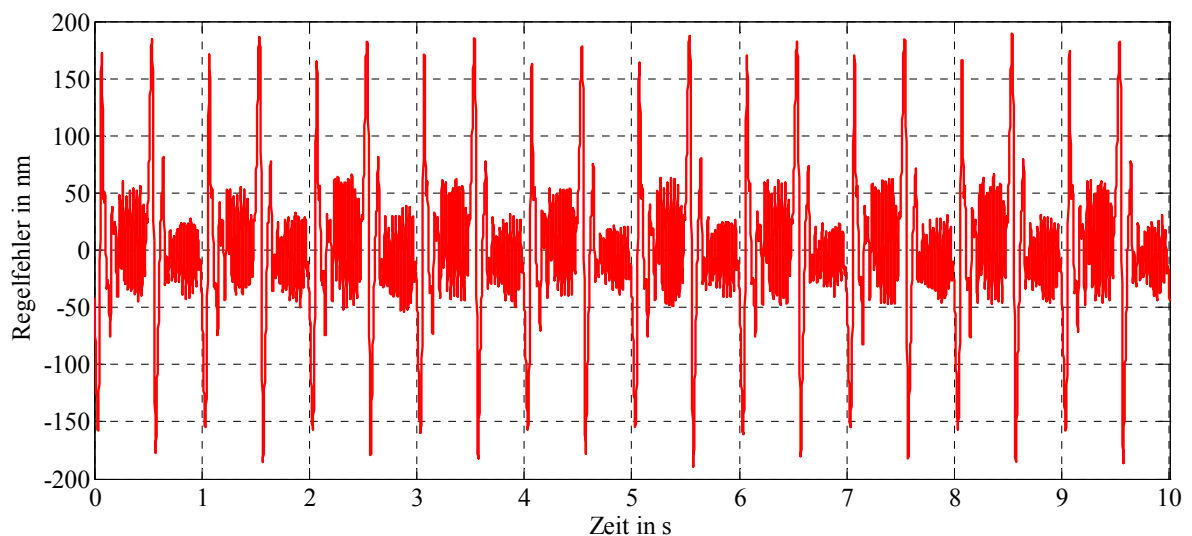


Abbildung A 29: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem KNN als Vorsteuerung und einem Störbeobachter

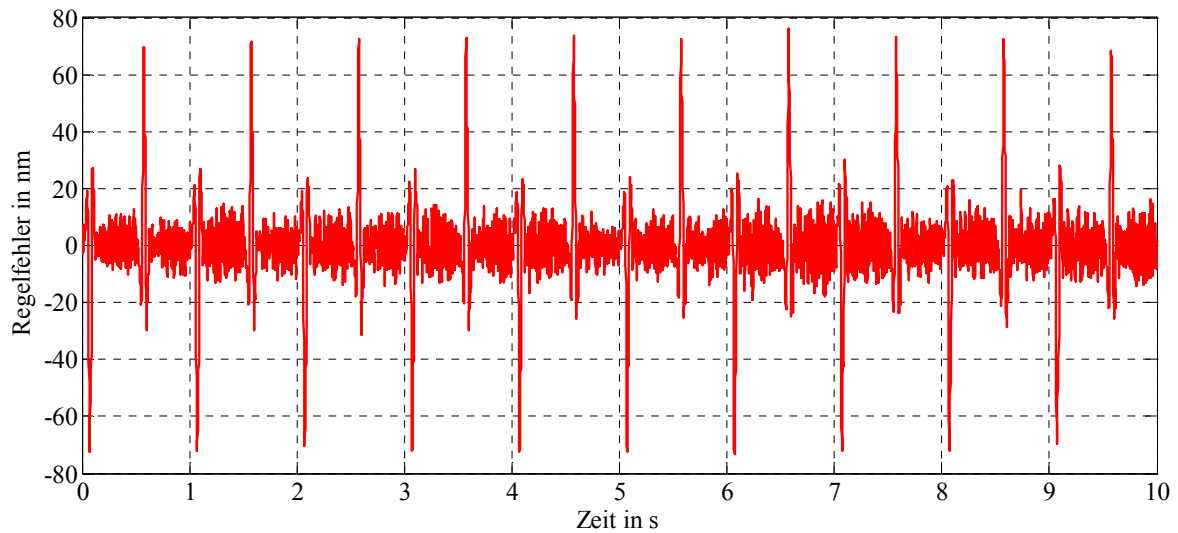


Abbildung A 30: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung und einem Störbeobachter

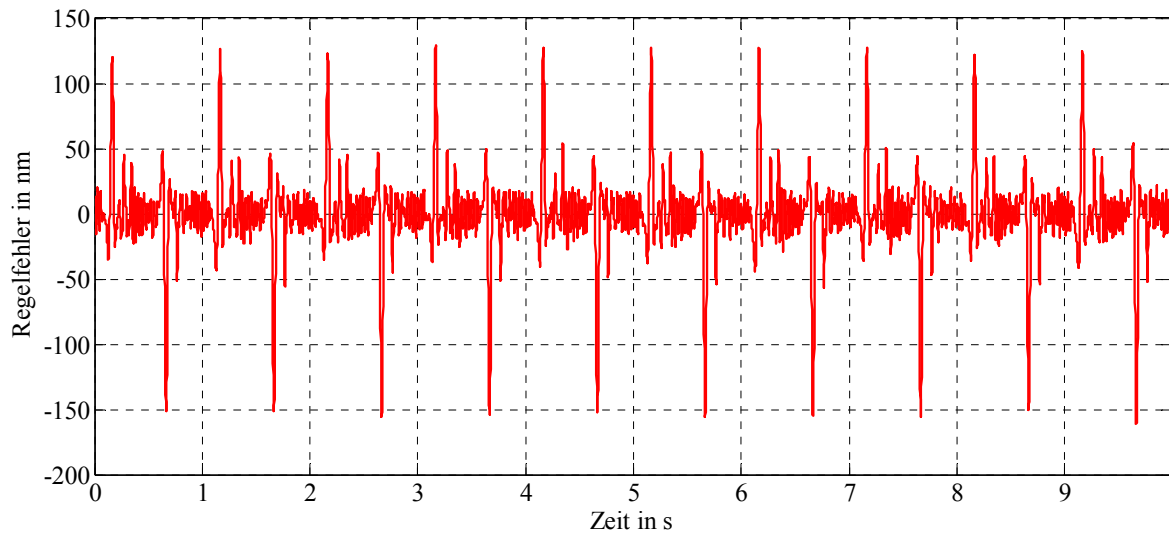


Abbildung A 31: Regelfehler bei einer sinusförmigen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem KNN als Vorsteuerung und einem Störbeobachter

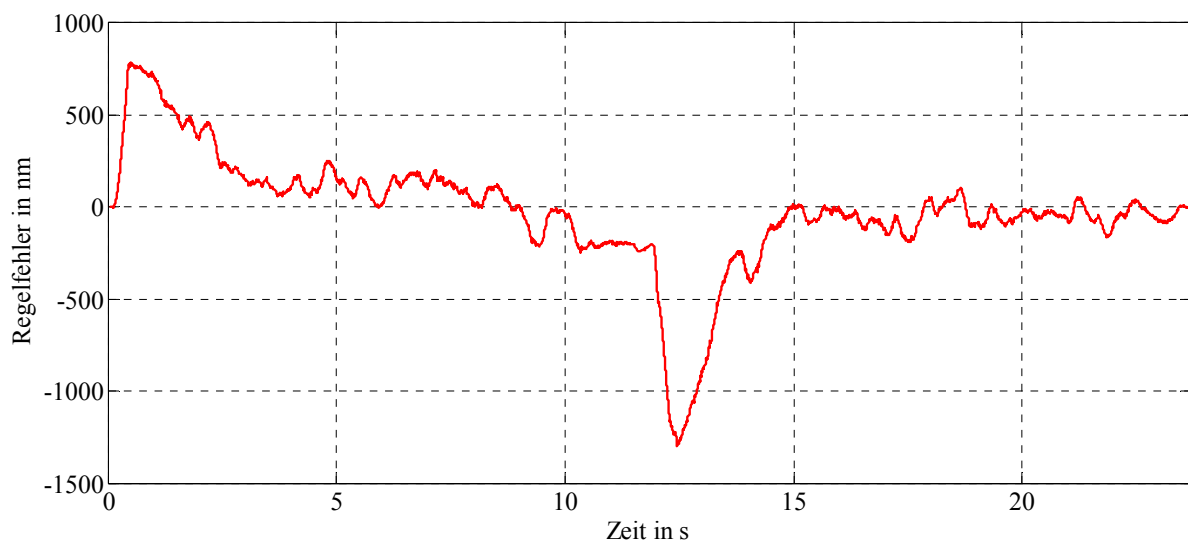


Abbildung A 32: Regelfehler bei einer linearen Bewegung unter Verwendung eines klassischen PID-Reglers

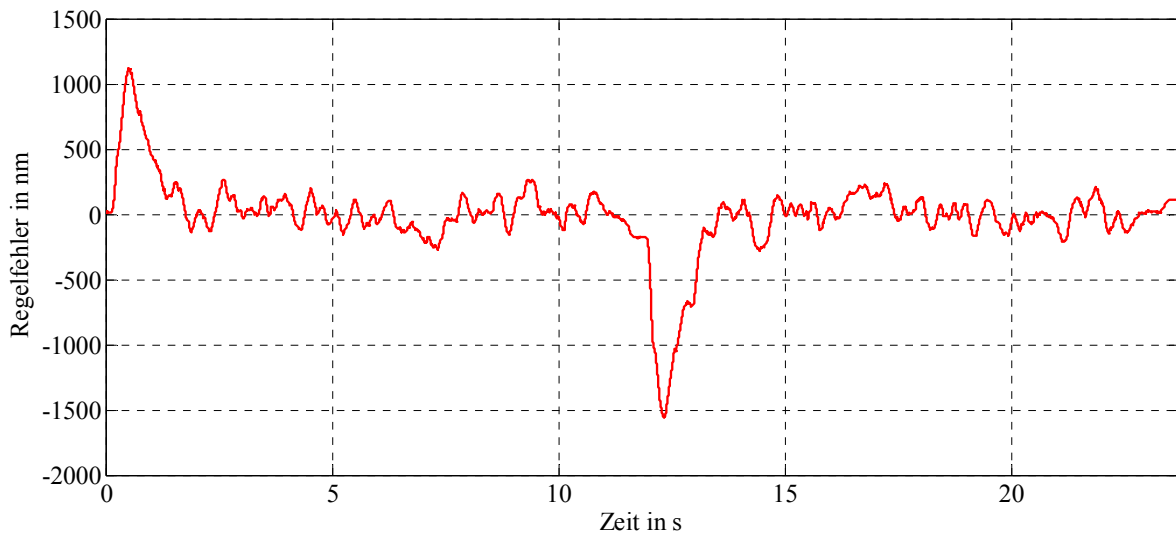


Abbildung A 33: Regelfehler bei einer linearen Bewegung unter Verwendung eines PI-Zustandsreglers

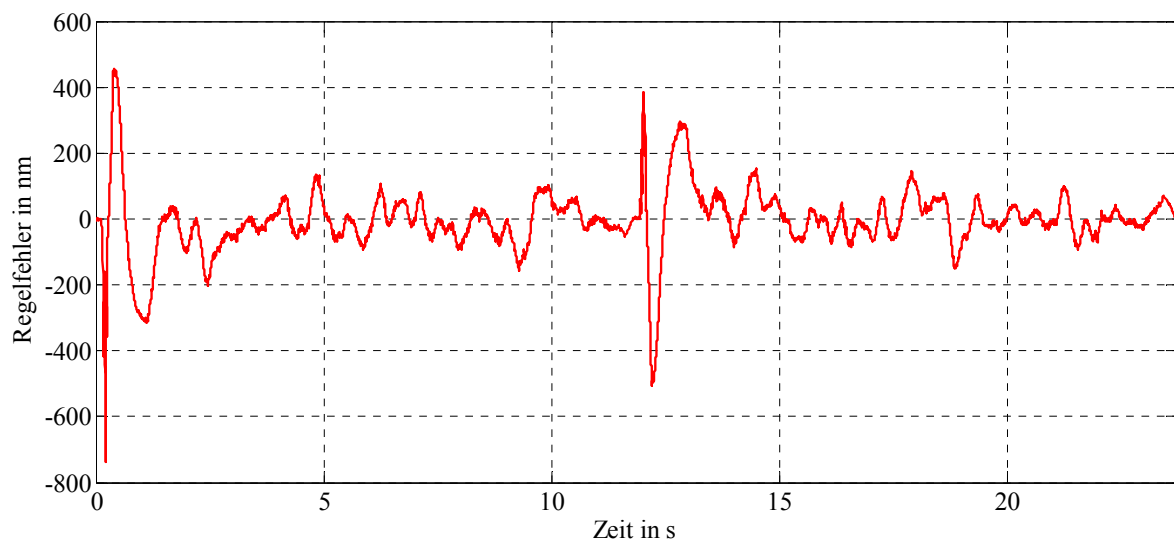


Abbildung A 34: Regelfehler bei einer linearen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung

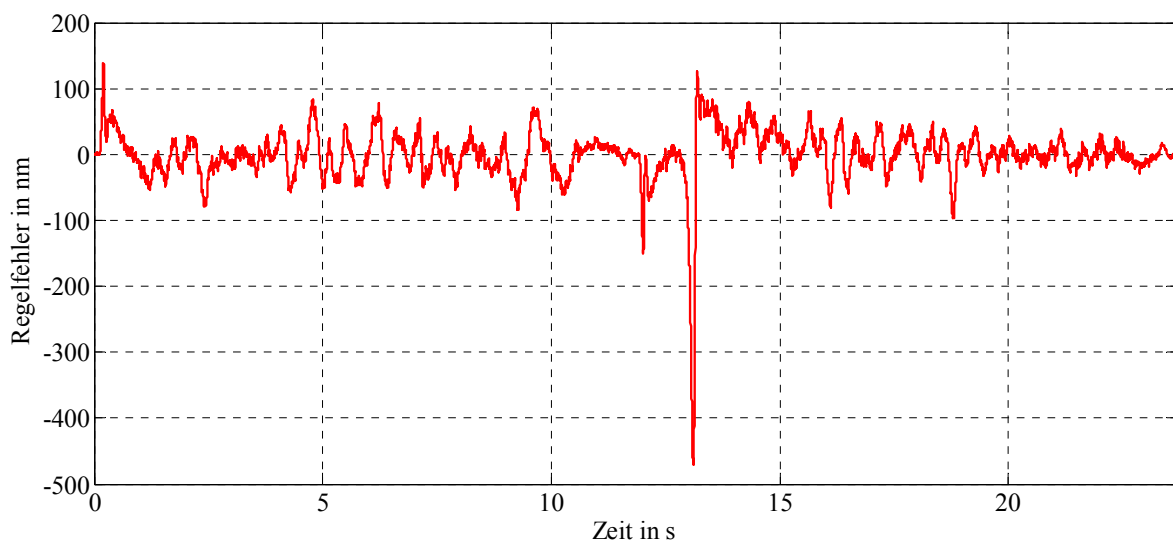


Abbildung A 35: Regelfehler bei einer linearen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem KNN als Vorsteuerung

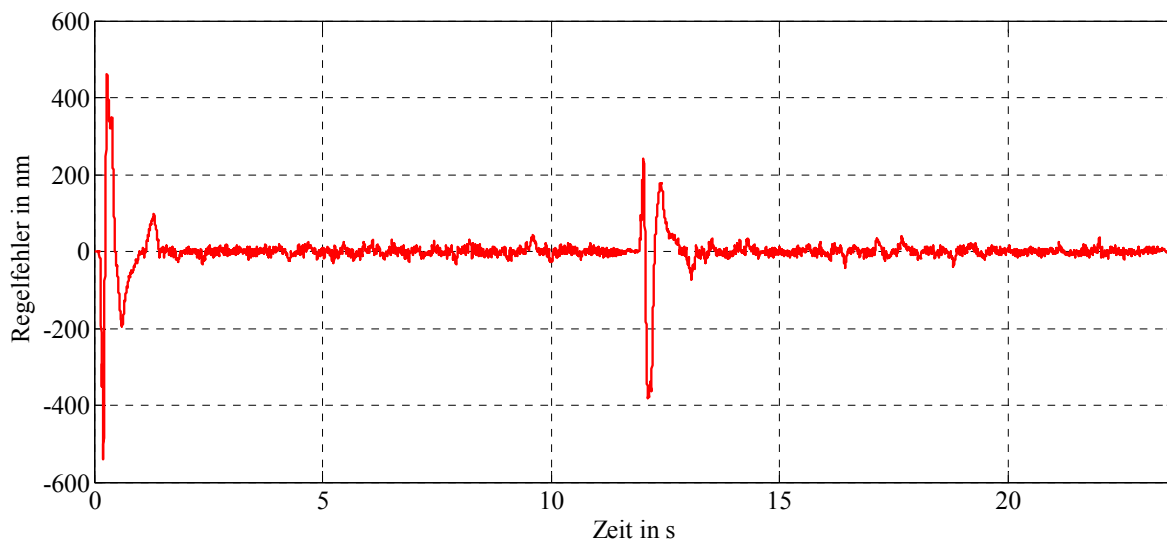


Abbildung A 36: Regelfehler bei einer linearen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung

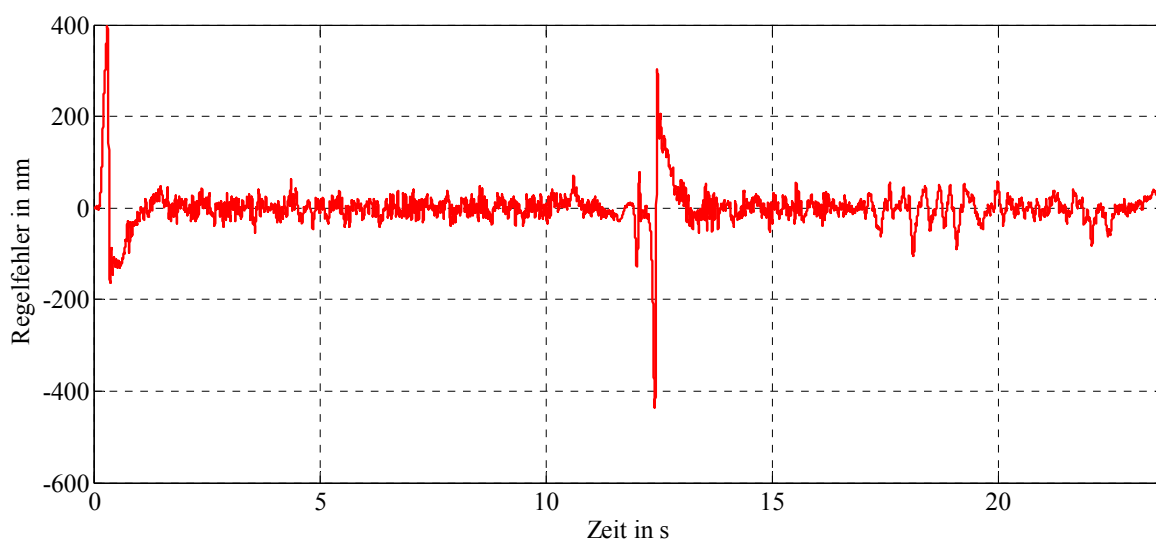


Abbildung A 37: Regelfehler bei einer linearen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem KNN als Vorsteuerung

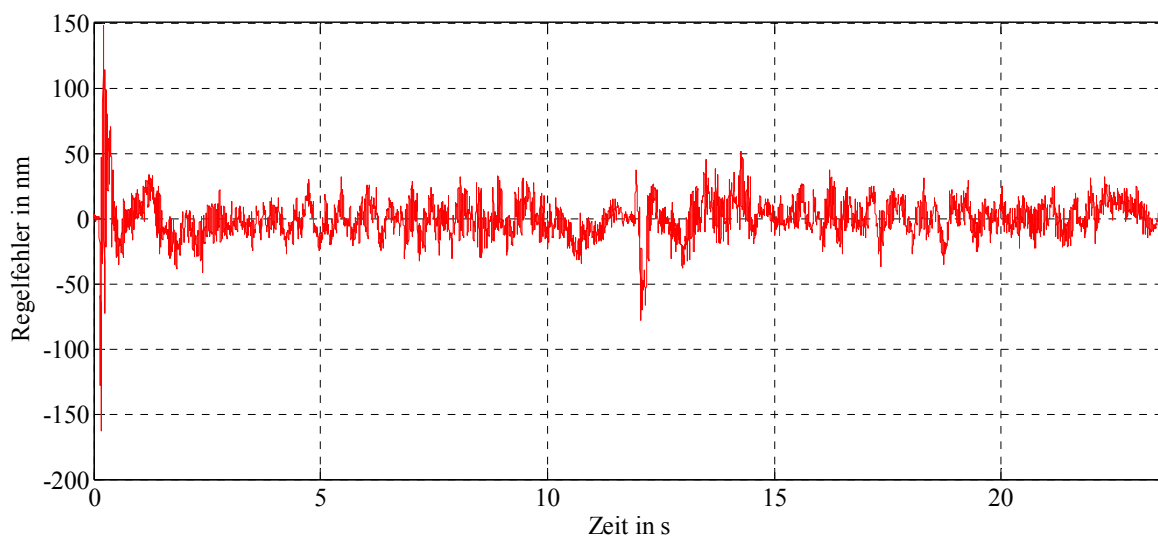


Abbildung A 38: Regelfehler bei einer linearen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung und einem Störbeobachter

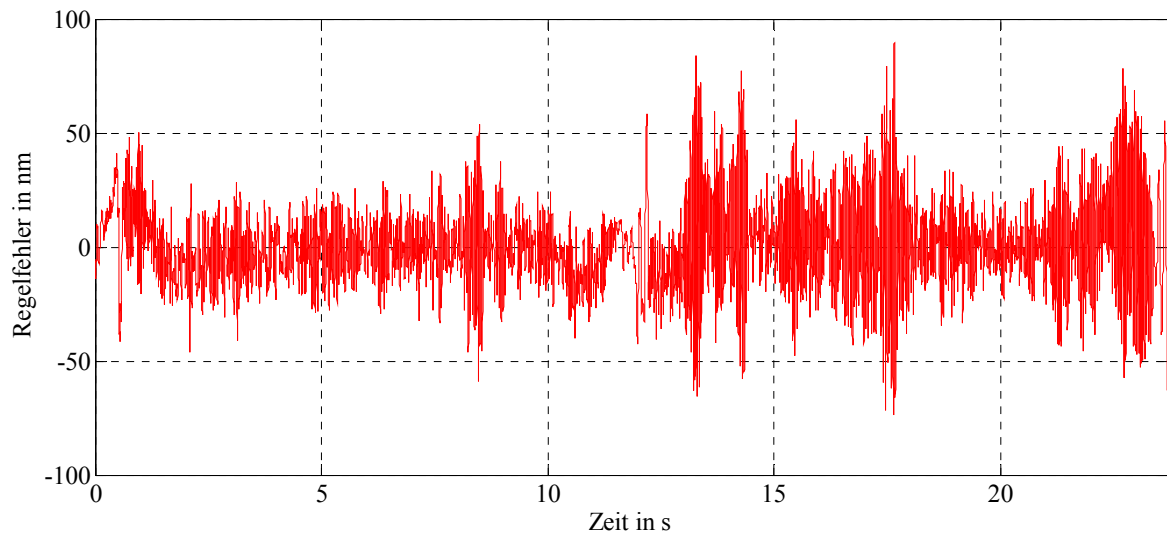


Abbildung A 39: Regelfehler bei einer linearen Bewegung unter Verwendung eines klassischen PID-Reglers in Kombination mit dem KNN als Vorsteuerung und einem Störbeobachter

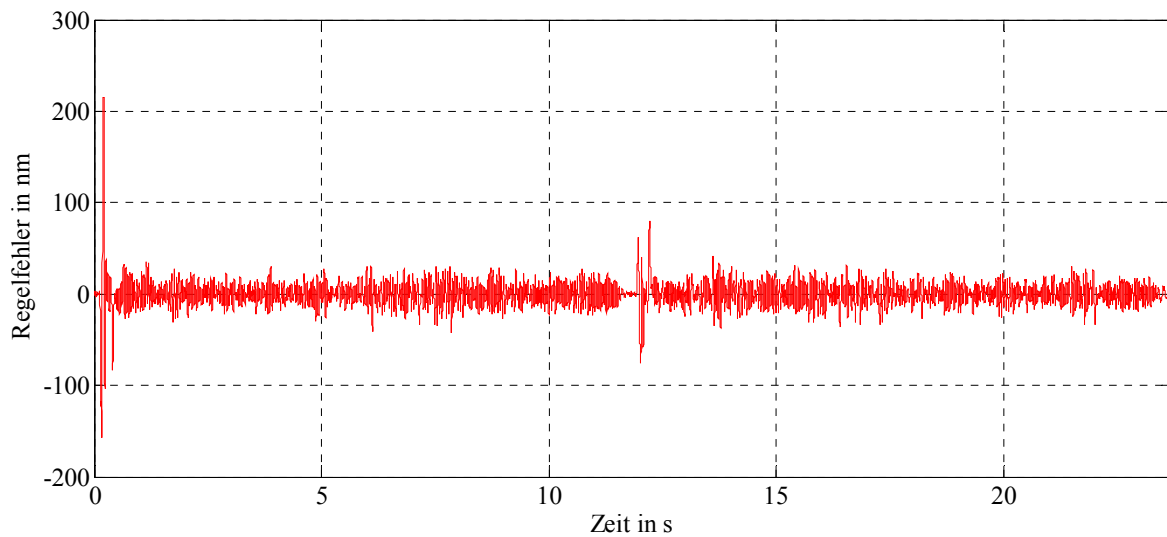


Abbildung A 40: Regelfehler bei einer linearen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem ADNLRX-Reibmodell als Vorsteuerung und einem Störbeobachter

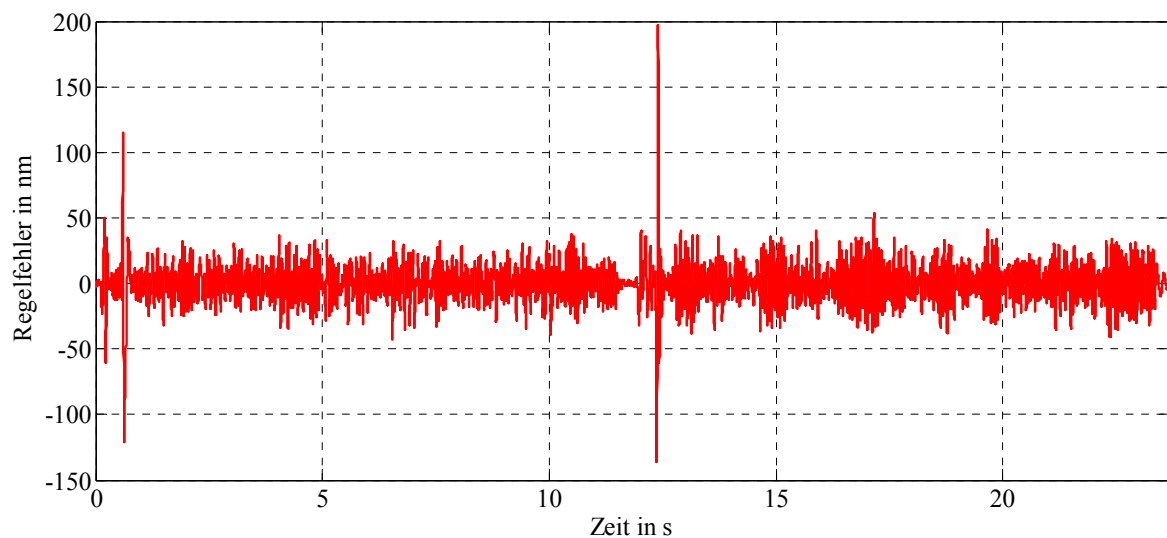


Abbildung A 41: Regelfehler bei einer linearen Bewegung unter Verwendung eines PI-Zustandsreglers in Kombination mit dem KNN als Vorsteuerung und einem Störbeobachter

Anhang D

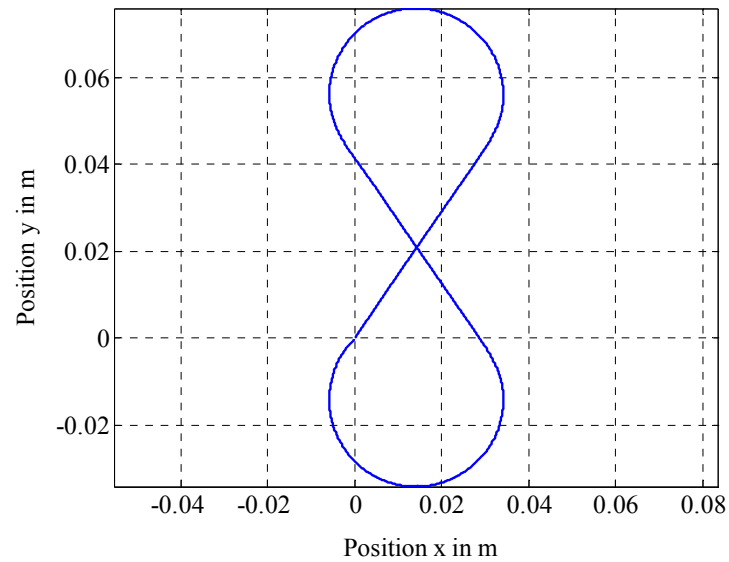


Abbildung A 42: Zweidimensionale Figur A (maximale Ausdehnung von 100x40 mm²)

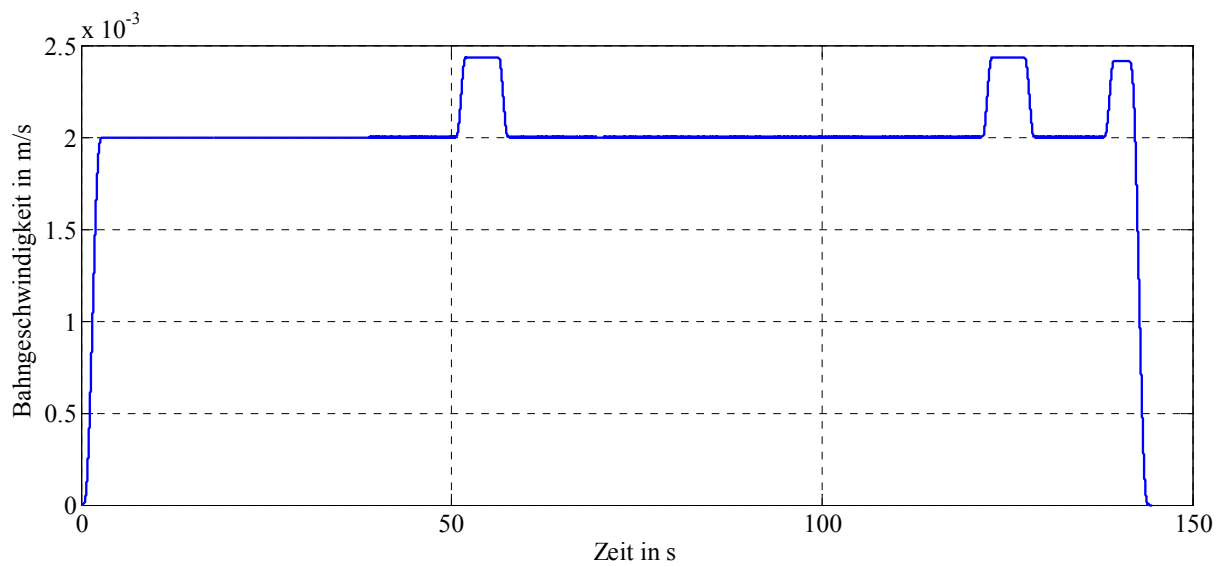


Abbildung A 43: Profil der Bahngeschwindigkeit von Figur A

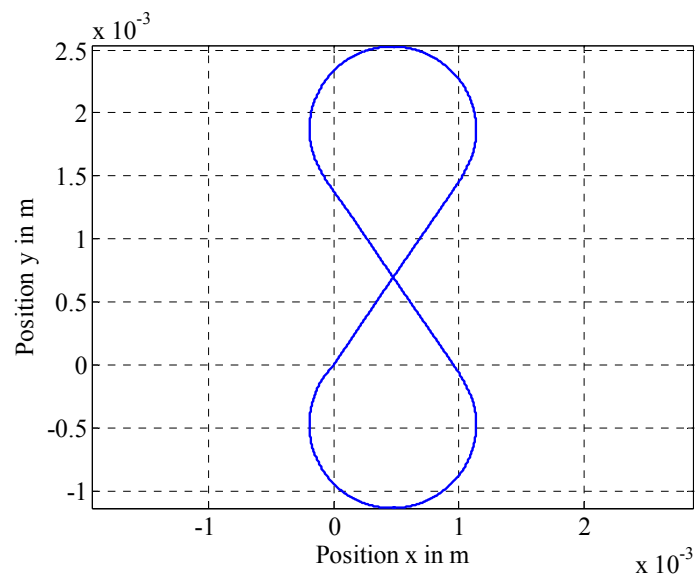


Abbildung A 44: Zweidimensionale Figur B (maximale Ausdehnung von $3,3 \times 1,3 \text{ mm}^2$)

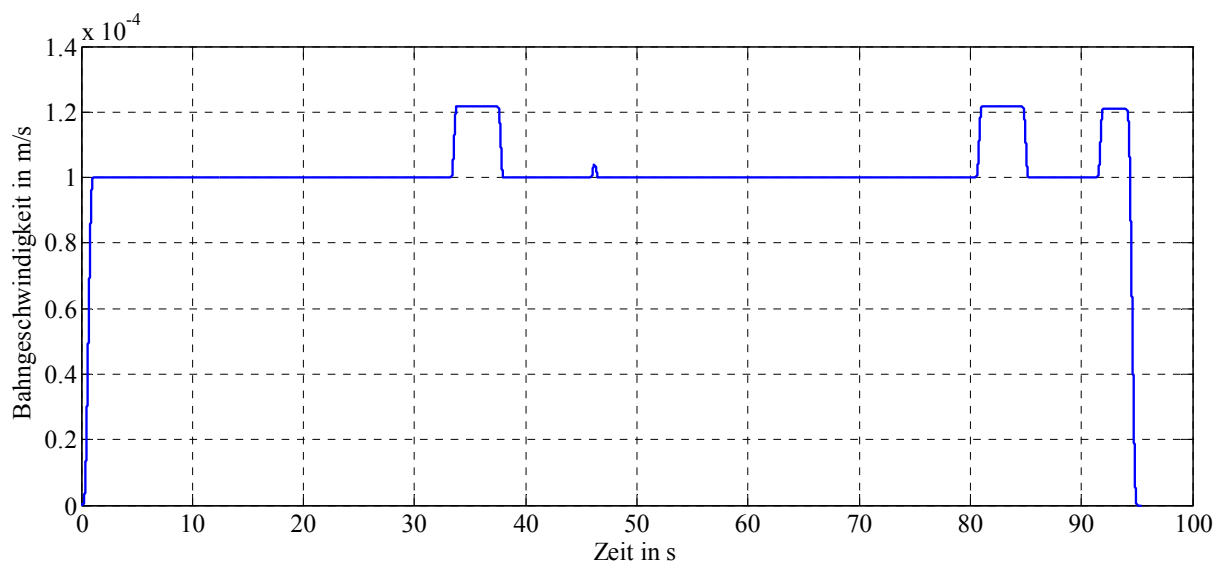


Abbildung A 45: Profil der Bahngeschwindigkeit von Figur B

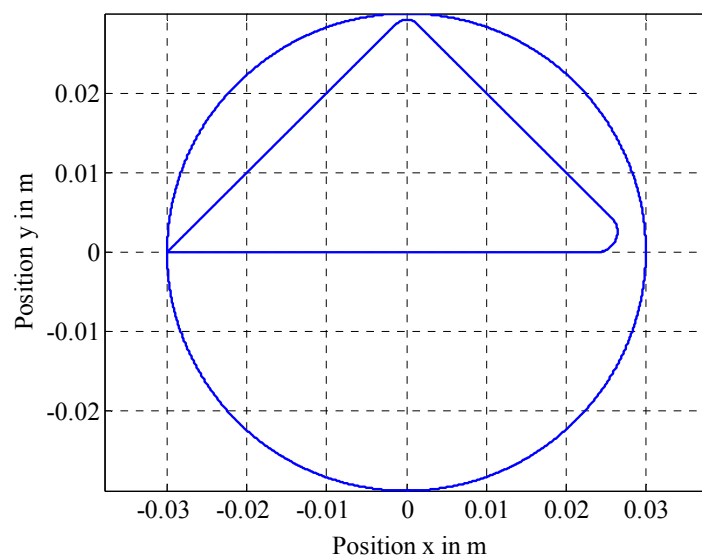


Abbildung A 46: Zweidimensionale Figur C (maximale Ausdehnung von $60 \times 60 \text{ mm}^2$)

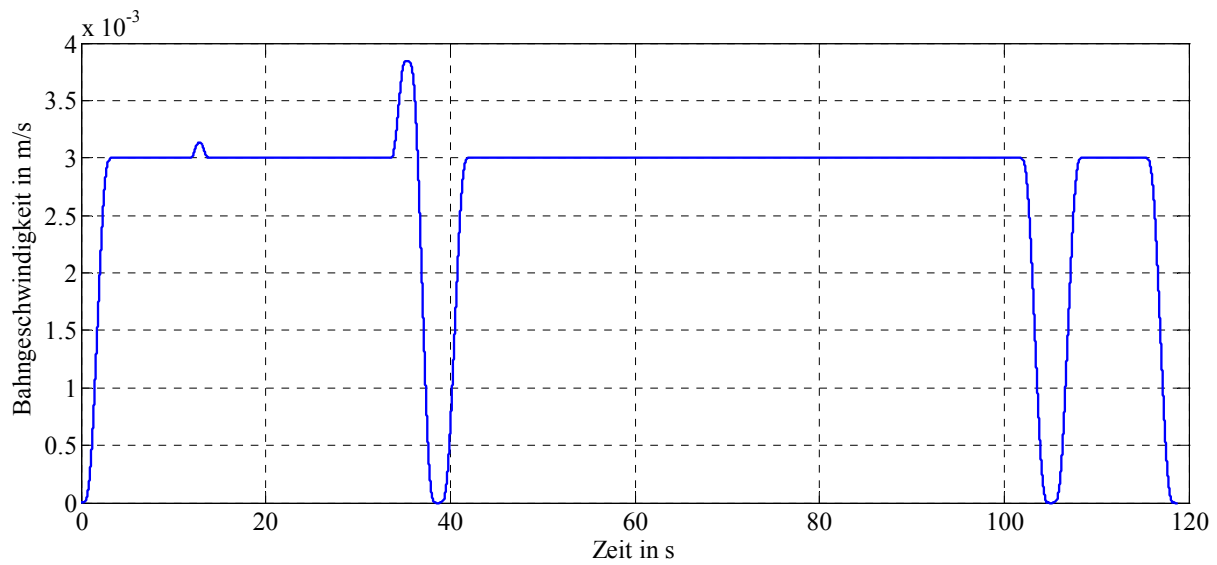


Abbildung A 47: Profil der Bahngeschwindigkeit von Figur C

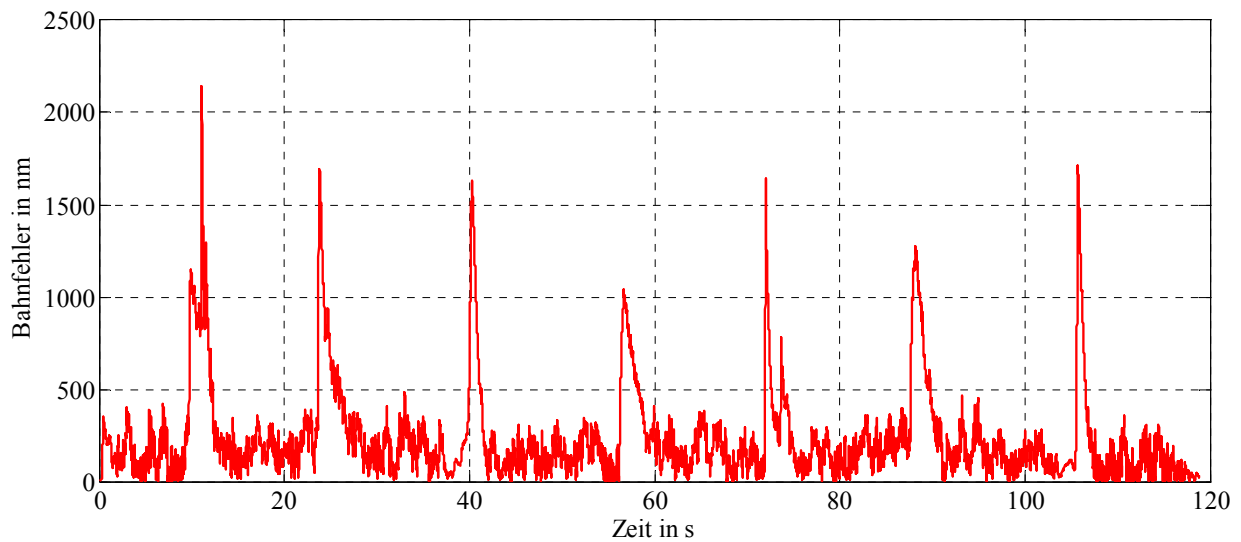


Abbildung A 48: Bahnabweichung beim Abfahren von Figur C unter Nutzung eines dezentralen PID-Reglers

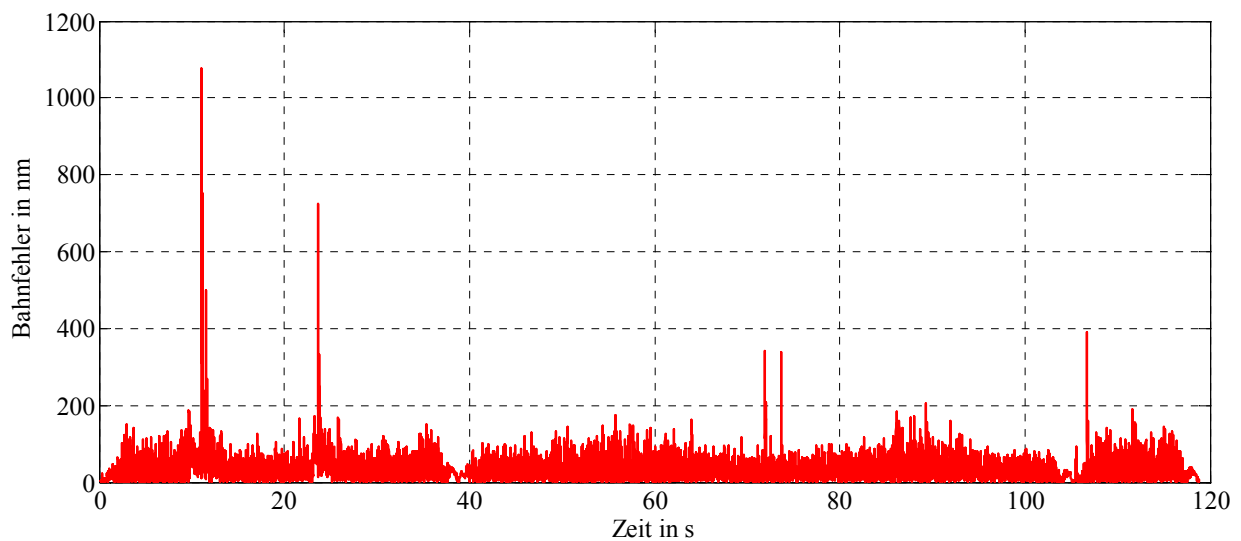


Abbildung A 49: Bahnabweichung beim Abfahren von Figur C unter Nutzung des dezentralen modellbasierten Regelungskonzeptes

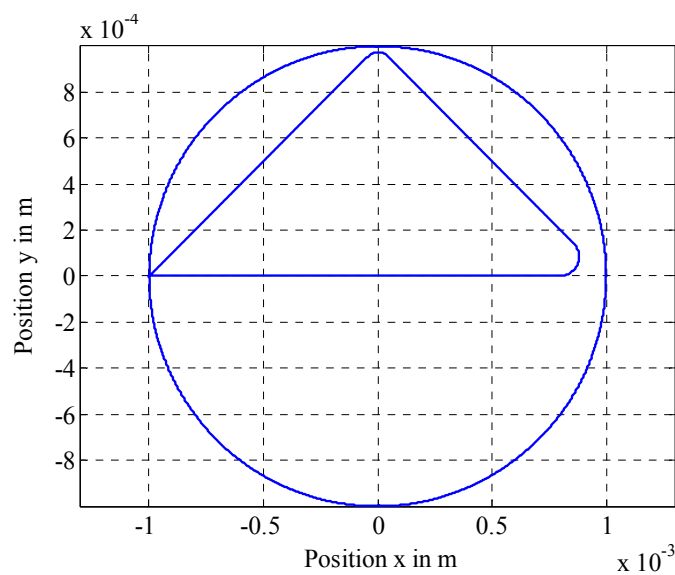


Abbildung A 50: Zweidimensionale Figur D (maximale Ausdehnung von $2 \times 2 \text{ mm}^2$)

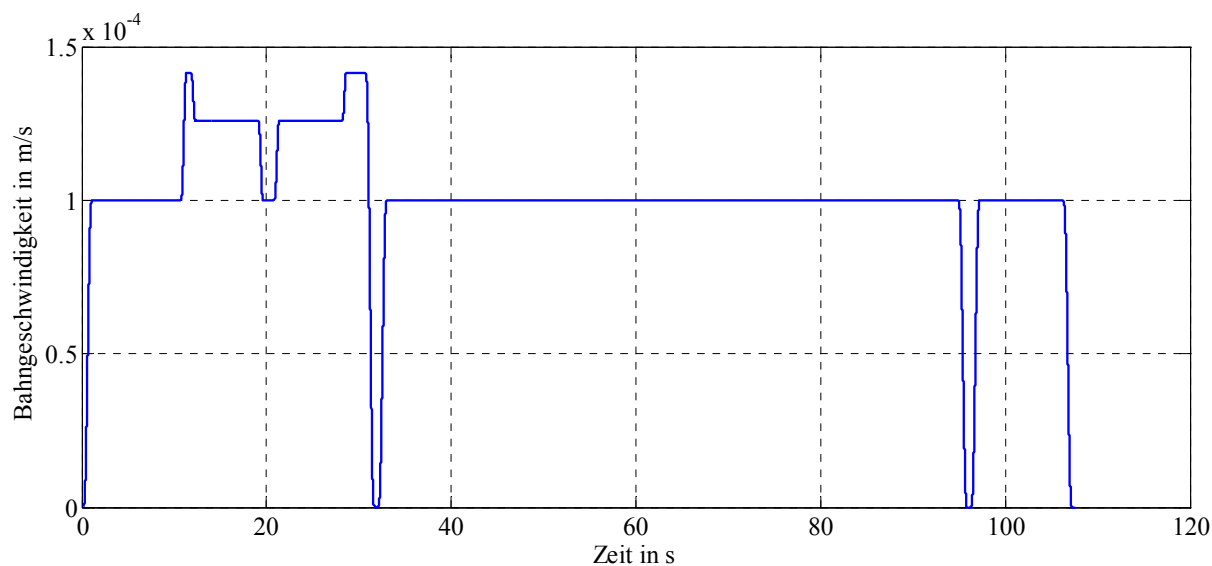


Abbildung A 51: Profil der Bahngeschwindigkeit von Figur D

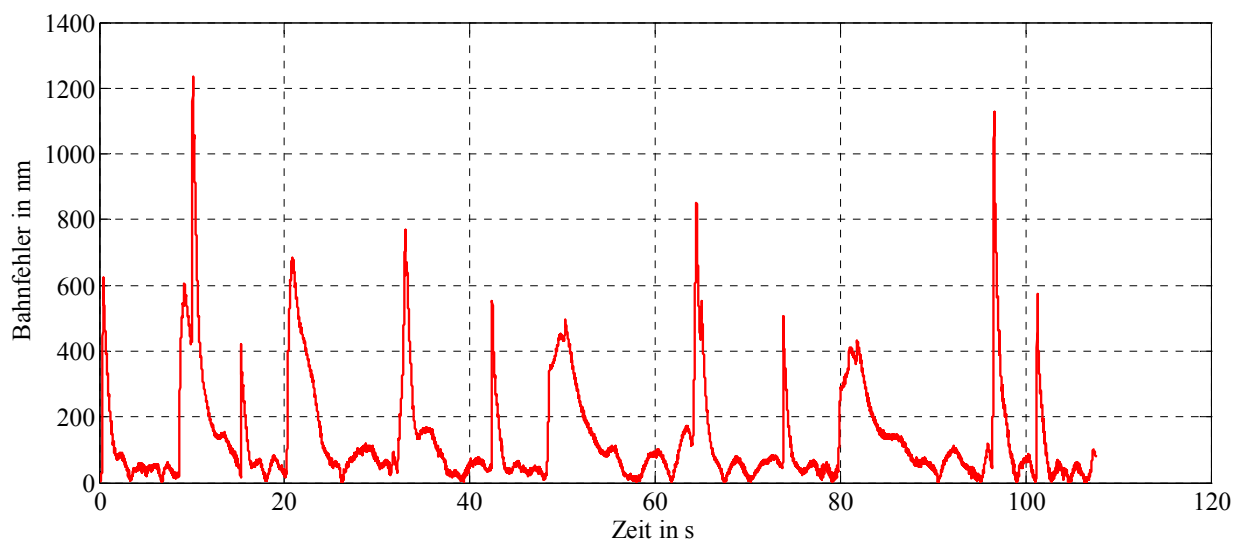


Abbildung A 52: Bahnabweichung beim Abfahren von Figur D unter Nutzung eines dezentralen PID-Reglers

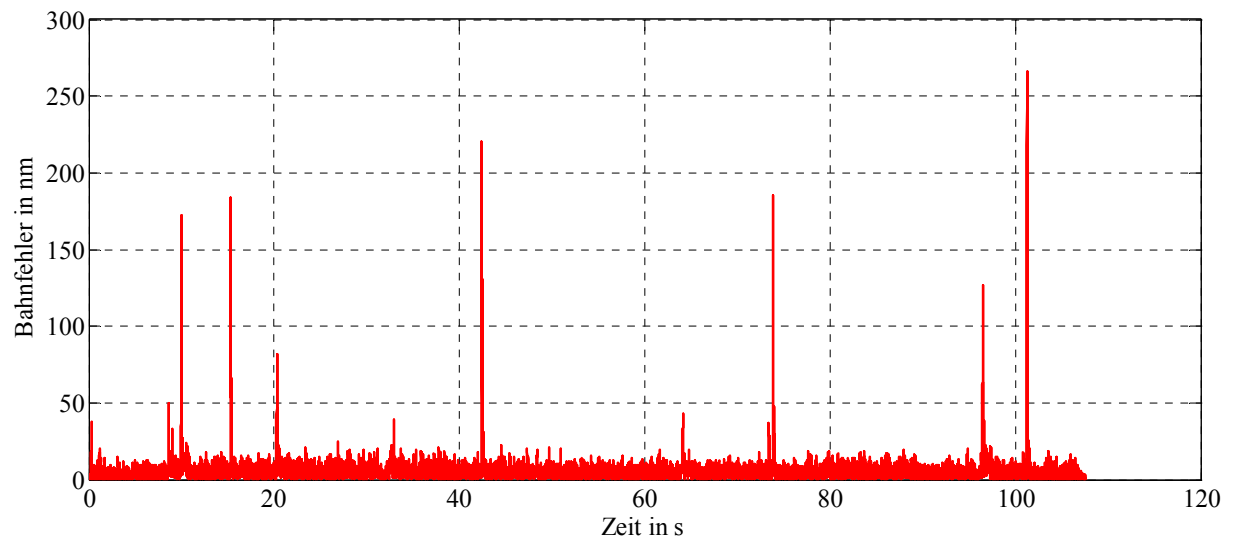


Abbildung A 53: Bahnabweichung beim Abfahren von Figur D unter Nutzung des dezentralen modellbasierten Regelungskonzeptes

Literaturverzeichnis

- [1] ABBE, E.: Meßapparate für Physiker., In: *Zeitschrift für Instrumentenkunde*, vol. 10, pp. 446-448, 1890
- [2] AL-BENDER, F. ; DE MOERLOOZE, K.: A Model of the Transient Behavior of Tractive Rolling Contacts., In: *Advances in Tribology*, vol. 2008, Article ID 214894, 17 Seiten, 2008
- [3] AL-BENDER, F. ; LAMPAERT, V. ; SWEVER, J.: The Generalized Maxwell-Slip Model: A novel Model for Friction Simulation and Compensation., In: *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1883-1887, 2005
- [4] AMONTONS, G.: De la resistance causée dans les machines., In: *Mémoires de l'Académie Royale*, pp. 206-226, 1699
- [5] AMTHOR, A. ; HAUSOTTE, T. ; AMENT, C. ; LI, P. ; JÄGER, G.: Friction Identification and Compensation on Nanometerscale., In: *IFAC World Congress 2008*, Proceedings, pp. 2014-2019, Seoul, 2008
- [6] AMTHOR, A. ; HAUSOTTE, T. ; LI, P. ; JÄGER, G.: Friction Modelling on Nanometerscale and Experimental Verification., In: *52. Internationales Wissenschaftliches Kolloquium*, Proceedings, vol. 1, pp. 495-501, Ilmenau, 2007
- [7] AMTHOR, A. ; WERNER, J. ; LORENZ, A. ; ZSCHAECK, S. ; AMENT, C.: Asymmetric motion profile planning for nanopositioning and nanomeasuring machines., In: *Proc. IMechE, Part I: J. Systems and Control Engineering*, 2010, 224 (1), 79-92. DOI 10.1243/09596518JSCE826
- [8] AMTHOR, A. ; ZSCHAECK, S. ; AMENT, C.: Position Control on Nanometer Scale based on an Adaptive Friction Compensation Scheme., In: *34th Annual Conference of the IEEE Industrial Electronics Society*, Proceedings, pp. 2568-2573, Orlando (USA), 2008
- [9] AMTHOR, A. ; ZSCHAECK, S. ; AMENT, C.: Adaptive Reibkraftkompensation zur modellbasierten Positionsregelung von Nanopositionier- und Nanomeßmaschinen., In: *at-Automatisierungstechnik* vol. 59, no. 2, pp. 51-59, 2009
- [10] AMTHOR, A. ; ZSCHAECK, S. ; AMENT, C.: High Precision Position Control using an Adaptive Friction Compensation Approach., In: *IEEE Transactions on Automatic Control*, vol. 55, no. 1, pp. 274-278, 2010

- [11] ARMSTRONG-HÉLOUVRY, B.: Control of Machines with Friction., Springer Verlag, 1. Auflage, 1991
- [12] ARMSTRONG-HÉLOUVRY, B. ; DUPONT, P. ; DE WIT, C.: A survey of Models, Analysis Tools and Compensation Methods for Control of Machines with Friction., In: *Automatica*, vol. 30, no. 7, pp. 1083-1138, 1994
- [13] ARMSTRONG-HÉLOUVRY, B.: Frictional Lag and Stick-Slip., In: *IEEE International Conference on Robotics and Automation*, Proceedings, vol. 2, pp. 1448-1453, 1992
- [14] ARNS, R.G.: The other transistor: early history of the metal-oxidesemiconductor field-effect transistor., In: *Engineering Science and Education Journal*, vol. 7, no. 5, pp. 233-240, 1998
- [15] ARULAMPALAM, G. ; BOUZERDOUM, A.: A generalized feedforward neural network architecture for classification and regression., In: *Neural Networks*, vol. 16, no. 5-6, pp. 561-568, 2003
- [16] ASYLUM RESEARCH: Break the Nanocode with the Next Generation of AFM., <http://www.atomicforce.de/Bilder/Asylum-Research/CypherAFMBrochure.pdf>, 2009
- [17] ATTOCUBE SYSTEMS AG: Products and Solutions., <http://www.attocube.com/nanoSCOPY/overview.htm>, 2009
- [18] BÄR, C.: Elementare Differentialgeometrie., Gruyter Verlag, 1. Auflage, 2001
- [19] BASF AG: Europa dämmt grün Styrodur® C., http://www.basf.de/basf2/img/produkte/kunststoffe/styrodur/downloads/DE/Styrodur_Europa_daemmt_gruen.pdf?id=A*wh3EX_sbW229., 2009
- [20] BECKMANN, M.: Entwurf und Vergleich verschiedener Regelungsansätze für eine Positionsregelung für Nanopositionier- und Nanomessmaschinen., Diplomarbeit, Technische Universität Ilmenau, 2009
- [21] BEI TECHNOLOGIES, INC. - KIMCO MAGNETICS DIVISION: Voice Coil Actuators - An Application Guide., <http://www.beikimco.com/pdf/VCA%20App%20Product%20Guide.pdf>, 2009
- [22] BIERMAN, G.J.: Factorization Methods for Discrete Sequential Estimation., Academic Press New York, 1977
- [23] BIERMAN, G.J.: Measurement updating using the U-D factorization., In: *IEEE Conference on Decision and Control*, Proceedings, Houston, USA, 1975
- [24] BINNING, G. ; QUATE, C.F. ; GERBER, C.: Atomic Force Microscope., In: *Physical Review Letters*, vol. 56, no. 9, pp. 930-933, 1986

- [25] BINNING, G. ; ROHRER, H. ; GERBER, C. ; WEIBEL, E.: Surface Studies by Scanning Tunnelling Microscopy, In: *Physical Review Letters*, vol. 49, no. 1, pp. 57-61, 1982
- [26] BRAND, U. ; KIRCHHOFF, J.: A micro-CMM with metrology frame for low uncertainty measurements., In: *Measurement Science and Technology*, vol. 16, pp. 2489-2497, 2005
- [27] BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG: Technologieanalyse: Potentiale der Nanotechnologie., <http://www.bmbf.de/de/4877.php>, 2009
- [28] CARL ZEISS INDUSTRIELLE MESSTECHNIK GmbH: F25 Messen im Nanometerbereich., <http://www.zeiss.de/c12567b6005445a1/Contents-Frame/137b68b760f8115cc1256e6f003bdc6e>, 2009
- [29] COLEMAN, T.F. ; LI, Y.: An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds., In: *SIAM Journal on Optimization*, vol. 6, pp. 418-445, 1996.
- [30] DA VINCI, L.: The Notebooks of Leonardo Da Vinci., Dover Publications Inc., 1970
- [31] DAHL, P.: A solid friction model., The Aerospace Corporation, El-Segundo, California, USA, 1968
- [32] DE MOERLOOZE, K. ; AL-BENDER, F.: Experimental Investigation into the Tractive Prerolling Behaviour of Balls in V-Grooved Tracks., In: *Advances in Tribology*, vol. 2008, Article ID 561280, 10 Seiten, 2008
- [33] DE WIT, C.C. ; OLSSON, H. ; ASTÖM, K.J. ; LISCHINSKY, P.: A New Model for Control of Systems with Friction., In: *IEEE Transactions on Automatic Control*, vol. 40, no. 3, pp. 419-425, 1995
- [34] DE WIT, C.C. ; SICILIANO, B. ; BSTIN, G.: Theory of Robot Control., Springer, Berlin, 1996
- [35] DE WIT, C.C.: Comments on "A New Model for Control of Systems with Friction"., In: *IEEE Transactions on Automatic Control*, vol. 43, no. 8, pp. 1189-1190, 1998
- [36] DESVASIA, S. ; ELEFThERIOU, E. ; MOHEIMANI, S.O.R.: A Survey of Control Issues in Nanopositioning., In: *IEEE Transactions on Control Systems Technology*, vol. 15, no. 5, 2007
- [37] DIN: Längenprüftechnik 3 – Messgeräte, Messverfahren., Beuth Verlag, 11. Auflage, 2008
- [38] DSPACE GmbH: Regler Prototyping., <http://www.dspace.de/ww/de/gmb/home/applicationfields/researcheducation/controlprototyping.cfm>, 2009

- [39] DUPONT, P. ; ARMSTRONG, B. ; HAYWARD, V.: Elasto-Plastic Friction Model: Contact Compliance and Stiction., In: *AACC American Control Conference*, Proceedings, pp. 1072-1077, Chicago (USA), 2000
- [40] DUPONT, P. ; HAYWARD, V. ; ARMSTRONG, B. ; ALTPETER, F.: Single State Elasto-Plastic Friction Models., In: *IEEE Transactions on Automatic Control*, vol. 47, no. 5, pp. 787 – 792, 2002
- [41] ERKORKMAZ, K. ; ALTINATZ, Y.: High speed CNC system design. Part I: jerk limited trajectory generation and quintic spline interpolation., In: *International Journal of Machine Tools & Manufacture*, vol. 41, no. 9, pp. 1323-1345, 2001
- [42] FAN, K.C. ; FEI, J.T. ; YU, X.F. ; CHEN, Y.J. ; WANG, W.L. ; CHEN, F. ; LIU, Y.S.: Development of a low-cost micro-CMM for 3D micro/nano measurements., In: *Measurement Science and Technology*, vol. 17, pp. 524-532, 2006
- [43] FÖLLINGER, O.: Regelungstechnik., Huethig Verlag Heidelberg, 8. Auflage, 1992
- [44] GERHARDT, U.: Signalverarbeitung in der interferometrischen Meß- und Sensortechnik., Dissertation, Technische Universität Ilmenau, 1997
- [45] GERTHSEN, CH. ; MESCHDE, D.: Gerthsen Physik., Springer Verlag, 2002
- [46] GRUHLKE, M.: Untersuchungen für den industriellen Einsatz der Nanometerkoordinatenmesstechnik, Ph.D. Thesis, Universität der Bundeswehr Hamburg, 2009
- [47] HAESSING, D. ; FRIEDLAND, B.: On the Modeling and Simulation of Friction., In: *ASME Journal of dyn. Systems, Measurement and Control*, vol. 113, no. 3, pp. 354-363, 1991
- [48] HAUSOTTE, T. ; JÄGER, G. ; MANSKE, E. ; SAWODNY, O.: Control System of a Nanopositioning and Nanomeasuring Machine., In: *9th International Conference on New Actuators*, Proceedings, pp. 123-125, Bremen, Germany, 2005
- [49] HAUSOTTE, T.: Nanopositionier- und Nanomessmaschine., Dissertation, Technische Universität Ilmenau, 2002
- [50] HESS, D.P. ; SOOM, A.: Friction at a lubricated line contact operating at oscillating sliding velocities., In: *Journal of Tribology*, vol. 112, no. 1 pp. 147-152, 1990
- [51] HOCKEN, R.J. ; FESPERMAN, R. ; OVERCASH, J. ; OZTURK, O. ; STROUP, C.: Engineering Nanotechnology: The Top Down Approach., In: *Key Engineering Materials*, vol. 381-382, pp. 3-6, 2008
- [52] HOLMES, M. ; HOCKEN, R.J. ; TRUMPER, D.: The Long-Range Scanning Stage: A Novel Platform for Scanned-Probe Microscopy., In: *Precision Engineering*, vol. 24, no. 3, pp. 191-209, 2000

- [53] HOUK, C.R. ; JOINES, J.A.; KAY M.G.: A Genetic Algorithm for Function Optimization: A Matlab Implementation., <http://www.ise.ncsu.edu/mirage/GAToolBox/gaot/>
- [54] HYUN, D. ; LANGARI, R.: Fuzzy Logic based Compensation of Friction in low Speed Motion Control., In: *IEEE International Symposium of Intelligent Control*, Proceedings, pp. 337-342, 2001
- [55] IA.CMM- THE ASSOCIATION OF CMM MANUFACTURES: I++ DME Dimensional Measurement Equipment Interface., [http://resources.renishaw.com/download/\(d5ebdfc1cb824d449036033a28e8430d\)?lang=en&inline=true](http://resources.renishaw.com/download/(d5ebdfc1cb824d449036033a28e8430d)?lang=en&inline=true), 2009
- [56] IDAM, INA - DRIVES & MECHATRONICS GMBH & Co. oHG., <http://www.inadam.de>
- [57] IMMS (INSTITUT FÜR MIKROELEKTRONIK- UND MECHATRONIK-SYSTEME) GMBH: Forschung und Entwicklung in Mikroelektronik und Mechatronik., <http://www.imms.de>, 2009
- [58] INTEL CORPORATION: Intel, Samsung Electronics und TSMC vereinbaren Übergang zu 450 mm Wafer Herstellung., <http://www.intel.com/cd/corporate/pressroom/emea/deu/archive/2008/391579.htm>, 2008
- [59] INTEL CORPORATION: Warum Mikroprozessoren immer kleiner werden., http://www.intel.com/corporate/pressroom/emea/deu/corei7/pdfs/Hintergrundinformation_Strukturgrößen.pdf, 2009
- [60] INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS: ITRS Public Conference, http://www.itrs.net/Links/2007Winter/2007_Winter_Presentations/01_ORTC_2007_JP.pdf, 2007
- [61] ISERMANN, R.: Identifikation dynamische Systeme., Springer Verlag Berlin, 2. Auflage, 1992
- [62] JÄGER, G. ; MANSKE, E. ; FÜSSL, R. ; GRÜNWALD, R. ; BÜCHNER, H.-J.: Nanometer-scale miniature interferometers developed for nanomeasurements in micro- and nanotechnology., In: *4th Euspen International Conference and 6th Annual General Meeting*, Proceedings, Glasgow, Scotland, 2004
- [63] JÄGER, G. ; MANSKE, E. ; HAUSOTTE, T. ; BÜCHNER, H.-J.: Nanomeßmaschinen zur abbefehlerfreien Koordinatenmessung., In: *Technisches Messen*, vol. 67, no. 7-8, pp. 319-323, 2000
- [64] JÄGER, G. ; MANSKE, E. ; HAUSOTTE, T. ; SCHOTT, W.: Operation and Analysis of a Nanopositioning and Nanomeasuring Machine., In: *Annual Meeting of the American Society for Precision Engineering*, Proceedings, pp. 299-304, St.Louis, USA, 2002

- [65] JÄGER, G. ; MANSKE, E. ; HAUSOTTE, T.; FÜSSL, R. ; GRÜNWALD, R. ; BÜCHNER, H.-J.: Optical Fibre Coupled Miniature Interferometers Designed for Application in Micro and Nano Devices., In: *19th Annual Meeting of the ASPE*, Orlando, USA, Proceedings, pp. 145 – 148, 2004
- [66] KALMAN, R.E.: A New Approach to Linear Filtering and Prediction Problems., In: *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, pp. 35-45, 1960
- [67] KARNOPP, D.: Computer simulation of stick-slip friction in mechanical dynamic systems., In: *ASME Journal of dyn. Systems, Measurement and Control*, vol. 7, no. 1, pp. 100-103, 1985
- [68] KIESSHAUER, C.: Reproduzierbare Positionierung mit Subnanometerauflösung in zwei räumlichen Achsen., Diplomarbeit, Technische Universität Ilmenau, 2009
- [69] KÜNG, A. ; MELI, F. ; THALMANN, R.: Ultraprecision micro-CMM using a low force 3D touch probe., In: *Measurement Science and Technology*, vol. 18, pp. 319-327, 2007
- [70] LAMBRECHTS, P. ; BOERLAGE, M. ; STEINBUCH, M.: Trajectory Planning and Feedforward Design for High Performance Motion Systems., In: *American Control Conference*, Proceedings, pp. 4637-4642, Boston (USA), 2004
- [71] LAMBRECHTS, P. ; BOERLAGE, M. ; STEINBUCH, M.: Trajectory planning and feedforward design for electromechanical motion systems., In: *Control Engineering Practice*, vol. 13, no. 3, pp. 145-157, 2004
- [72] LAMPAERT, V. ; AL-BENDER, F. ; SWEVERS, J.: A Generalized Maxwell-Slip Friction Model appropriate for Control Purposes., In: *IEEE International Conference for Physics and Control*, Proceedings, pp. 1170-1177, St. Petersburg (Russia), 2003
- [73] LAMPAERT, V. ; AL-BENDER, F. ; SWEVERS, J.: Experimental characterization of dry friction at low velocities on a developed tribometer setup for macroscopic measurements., In: *Tribology Letters*, vol. 16, no. 1–2, pp. 95-105, 2004
- [74] LAMPAERT, V. ; SWEVERS, J. ; AL-BENDER, F.: Experimental Comparison of different Friction Models for accurate low-velocity Tracking., In: *10th Mediterranean Conference on control and Automation*, Lisbon, Portugal, Proceedings, 2002
- [75] LAMPAERT, V. ; SWEVERS, J. ; AL-BENDER, F.: Modification of the Leuven Integrated Friction Model Structure., In: *IEEE Transactions on Automatic Control*, vol. 47, no. 4, pp. 683-687, 2002
- [76] LATOMBE, J.C.: Robot Motion Planning., Kluwer Academic Publications, 4. Ausgabe, Boston, 1996

- [77] LEACH, R. ; HAYCOCKS, J. ; JACKSON, K. ; LEWIS, A. ; OLDFIELD, S. ; YACOOT, A.: Advances in traceable nanometrology at the National Physical Laboratory., In: *Nanotechnology*, vol. 12, pp. R1-R6, 2001
- [78] LENZE, B.: Einführung in die Mathematik neuronaler Netze., Logos Verlag Berlin, 2. Auflage, 2003
- [79] LEONHARD, W.: Trajectory Control of a Multi-axes Robot with Electrical Servo Drives., In: *IEEE Conference on Industrial Electronics*, Proceedings, Philadelphia (USA), 1989
- [80] LERA, G. ; PINZOLAS, M.: Neighborhood Based Levenberg-Marquardt Algorithm for Neural Network Training., In: *IEEE Transactions on Neuronal Networks*, vol. 13, no. 5, pp. 1200-1203, 2002
- [81] LI, H.Z. ; GONG, Z.M. ; LIN, W ; LIPPA, T.: Motion profile planning for reduced jerk and vibration residuals., In: *SIMTech technical reports*, vol. 8, no. 1, pp. 32-37, 2007
- [82] LORENZ, A.: Entwicklung eines Bahnplanungsalgorithmus zur echtzeitfähigen Erzeugung von Referenztrajektorien unter Berücksichtigung von globalen kinematischen Beschränkungen., Diplomarbeit, Technische Universität Ilmenau, 2009
- [83] LUNZE, J.: Regelungstechnik 2: Mehrgrößensysteme, Digitale Regelung., Springer Verlag Berlin, 5. Auflage, 2008
- [84] LUTHER, W. ; MALANOWSKI, N.: Das wirtschaftliche Potential der Nanotechnologie., In: *Technikfolgenabschätzung – Theorie und Praxis*, vol. 13, no. 2, pp. 26-33, 2004
- [85] MAASS, W. ; NATSCHLAEGER, T. ; MARKRAM, H.: Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations., In: *Neural Computation*, vol. 14, pp. 2531-2560, 2002
- [86] MICOS GMBH: <http://www.micos-online.com/>, 2009
- [87] MOSTEFAI, L. ; DENAI, M. ; HORI, Y.: Fuzzy Observer-Based Control of Servomechanisms Subject to Friction Dynamics., In: *The 33rd Annual Conference of the IEEE Industrial Electronics Society*, Proceedings, pp. 328-332, 2007
- [88] MÜLLER, M. ; AMTHOR, A. ; FENGLER, W. ; AMENT, C: Model-driven Development and Multi-processor Implementation of a Dynamic Control Algorithm for Nanopositioning and Nanomeasuring Machines., In: *Journal of Systems and Control Engineering*, vol. 223, no. 3, pp. 417-429, 2009
- [89] OLOMSKI, J.: Trajectory Planning, Optimization and Control for Industrial Robots., In: *International Conference on Control and Applications*, Proceedings, Jerusalem, 1989

- [90] OLSSON, H. ; ASTÖM, K.J. ; LISCHINSKY, P ; DE WIT, C.C.: Friction Models and Friction Compensation., In: *European Journal of Control*, vol. 4, no. 3, pp. 176-195, 1998
- [91] PAPADOPOULOS, E.G. ; CHASPARIS, G.C.: Analysis and Model-based Control of Servomechanisms with Friction., In: *International Conference on Intelligent Robots and Systems (IROS 2002)*, EPFL Lausanne, Proceedings, 2002
- [92] PEGGS, G. ; LEWIS, A ; LEACH, R.: Measuring in three dimensions at mesoscopic level., In: *ASPE 2003 Winter Topical Meeting*, vol. 28, pp. 53-57, 2003
- [93] PERSSON, B.N.J.: Sliding Friction: Physical Principles and Applications., Springer Verlag, 2. Auflage, 2000
- [94] PIAZZI, A. ; VISIOLI, A.: Global Minimum-Jerk Trajectory Planning of Robot Manipulators., In: *IEEE Transactions on Industrial Electronics*, vol. 47, no. 1, pp. 140-149, 2000
- [95] POPOV, V.L.: Kontaktmechanik und Reibung. Ein Lehr- und Anwendungsbuch von der Nanotribologie bis zur numerischen Simulation., Springer-Verlag, 1. Auflage, 2009
- [96] RAMASUBRAMANIAM, A. ; RAY, L.R.: Friction Cancellation in Flexible Systems Using Extended Kalman-Bucy Filtering., In: *American Control Conference*, Proceedings, vol. 2, pp. 1062-1067, Denver, USA, 2003
- [97] RAMASUBRAMANIAM, A. ; RAY, L.R.: Stability and Performance Analysis of Non-Model-Based Friction Estimators., In: *Conference on Decision and Control*, Proceedings, vol. 3, pp. 2929-2935, Orlando, USA, 2001
- [98] RAY, L.R. ; RAMASUBRAMANIAM, A. ; TOWNSEND, J.: Adaptive Friction Compensation using Extended Kalman-Bucy filter friction estimation., In: *Control Engineering Practice*, vol. 9, no. 2, pp. 169-179, 2001
- [99] REGER, J. ; MAI, M. ; SIRA-RAMIREZ, H.-J.: Robust algebraic state estimation of chaotic systems., In: *IEEE Conference on Control Applications*, Proceedings, München, Deutschland, 2006
- [100] REGER, J. ; SIRA-RAMIREZ, H.-J. ; FLIESS, M.: On non-asymptotic observation of nonlinear systems., In: *IEEE Conference on Decision and Control*, Proceedings, Sevilla, Spanien, 2005
- [101] REYNOLDS, O.: On the theory of lubrications and its application to Mr. Beauchamp Tower's experiments, including an experimental determination of the viscosity of olive oil., In: *Philosophical Transactions of the Royal Society*, vol. 177, pp. 157-234, 1886

- [102] RIZOS, D.D. ; FASSOIS, D. ; SPLILIOS, D., Presliding Friction Identification Based Upon the Maxwell Slip Model Structure., In: *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 14, no. 2, pp. 431-445, 2004
- [103] RIZOS, D.D. ; FASSOIS, D.: Friction Identification based upon the LuGre and Maxwell Slip Models., In: *IEEE Transactions on Control Systems Technology*, vol. 17, no. 1, pp. 153-160, 2009
- [104] RIZOS, D.D. ; FASSOIS, D.: Maxwell Slip Model Based Identification and Control of Systems with Friction., In: *44th IEEE Conference on Decision and Control, and the European Control Conference 2005*, Sevilla, Spanien, Proceedings, 2005
- [105] RUIJI, T.: Ultra Precision Coordinate Measuring Machine; Design, Calibration, and Error Compensation., Ph.D. Thesis, TU Delft, 2001
- [106] SAWODNY, O ; ASCHEMANN, H. ; LAHRES, S. ; HOFER, E. P. A low cost material handling and logistic system for shop floor manufacturing using an automated bridge crane., In: *14th IFAC World Congress*, Proceedings, pp. 517-522, Beijing, 1999, Beijing, China
- [107] SAWODNY, O ; ASCHEMANN, H. ; LAHRES, S.: An automated gantry crane as a large workspace robot., In: *Control Engineering Practice*, vol. 10, no. 12, pp. 1323-1338, 2002
- [108] SAWODNY, O.: Kranautomatisierung und Großraumrobotik. Reihe 8, Nr. 888, Fortschritt-Berichte, VDI Verlag, Düsseldorf, 2001
- [109] SCHAUMSTOFFE HELGERS e.K.: <http://www.schalldaemmung-aixfoam.de/>, 2009
- [110] SCHMIDT, I.: Beiträge zur Verringerung der Positionierungsunsicherheit in der Nanopositionier- und Nanomessmaschine., Dissertation, Technische Universität Ilmenau, 2008
- [111] SCHNEEBERGER AG: Linearführungen und Umlaufkörper., http://www.schneebeger.com/downloadAction.de.de.do?content=6_1_1_ProductCatalogues¶m=cat_catalogues&reset=true, 2009
- [112] SCHOTT AG: Zerodur® Glaskeramik – Temperaturstabile Präzision und Qualität in Serie., http://www.schott.com/advanced_optics/german/download/zerodur_katalog_deutsch_2004.pdf, 2009
- [113] SCHOTT, W. ; DONTSOV, D. ; PÖSCHEL, W.: Mehrkanalige laserinterferometrische Messverfahren mit höchster Genauigkeit., In: *Photonik*, vol. 3, pp. 76-78, 2008
- [114] SELMIC, R.R. ; LEWIS, F.L.: Neural Network Approximation of Piecewise Continuous Functions: Application to Friction Compensation., In: *IEEE Transactions on neural networks*, vol. 13, no. 3, pp. 745-751, 2002
- [115] SIOS MESSTECHNIK GMBH: Miniaturinterferometer mit Planspiegelreflektor, Serie SP., http://www.sios.de/DEUTSCH/PRODUKTE/SP_D_FOT.PDF, 2009

- [116] SIOS MESSTECHNIK GMBH: Nanopositionier- und Nanomessmaschine., <http://www.sios.de/DEUTSCH/PRODUKTE/NMM.PDF>, 2009
- [117] SIOS MESSTECHNIK GMBH: Stabilisierter He-Ne-Laser., http://www.sios.de/DEUTSCH/PRODUKTE/SL_D_FOT.PDF, 2009
- [118] STÖLTING, H.-D. ; KALLENBACH, E.: Handbuch Elektrische Kleinantriebe., Hanser Verlag, 3. Auflage, 2006
- [119] STRIBECK, R.: Die wesentlichen Eigenschaften der Gleit- und Rollenlager., In: *Verein deutscher Ingenieure*, vol. 46, no. 38, 39, pp. 1342-1348/1432-1437, 1902
- [120] SWEVERS, J. ; AL-BENDER, F. ; GANSEMAN, C.G. ; PRAJOGO, T.: An Integrated Friction Model Structure with Improved Presliding Behaviour for Accurate Friction Compensation., In: *IEEE Transactions on Automatic Control*, vol. 45, no. 4, pp.675-686, 2000
- [121] TAKEUCHI, H. ; YOSIZUMI, K. ; TSUTSUMI, H.: Ultrahigh Accurate 3-D Profilometer., In: *Proceedings of the SPIE Optical Design and Testing II*, vol. 5638, pp. 387-394, 2005
- [122] TECHNISCHE UNIVERSITÄT ILMENAU: DFG-Sonderforschungsbereich 622., <http://www.tu-ilmenau.de/uni/DFG-Sonderforschun.1497.0.html>, 2009
- [123] THE MATHWORKS, INC.: Produkte und Dienstleistungen., <http://www.mathworks.de/products/index.html>, 2009
- [124] THE MATHWORKS, INC.: Real-Time Workshop® 7., <http://www.mathworks.com/mason/tag/proxy.html?dataid=9547&fileid=43808>, 2009
- [125] TIEN, S. ; ZOU, Q. ; DEVASIA, S.: Iterative Control of dynamics-Coupling Effects in Piezo Based Nano-Positioners for High Speed AFM., In: *IEEE Conference on Control Applications*, Taiwan, Proceedings, pp. 711-717, 2004
- [126] TJAHJOWIDODO, T. ; AL-BENDER, F ; VAN BRUSSEL, H.: Friction Identification and Compensation in a DC Motor., In: *IFAC World Congress*, Proceedings, Prag 2005
- [127] VAN SEGGLLEN, J.: NanoCMM., Ph.D. Thesis, TU Eindhoven, 2007
- [128] VEECO INSTRUMENTS, INC.: Metrology & Instrumentation Atomic Force Microscopes-Systems.,<http://www.veeco.com/atomic-force-microscope-systems/index.aspx>, 2009
- [129] WERNER, C.: Entwurf eines Regelungskonzeptes auf Basis von künstlichen Neuronalen Netzen., Studienarbeit, Technische Universität Ilmenau, 2008
- [130] WERNER, J.: Entwicklung und Implementierung von Bahnplanungsalgorithmen., Studienarbeit, Technische Universität Ilmenau, 2008

- [131] WERNER, J.: Weiterentwicklung und Implementierung eines Bahnplanungsalgorithmus zur echtzeitfähigen Erzeugung dynamischer Sollwerttrajektorien., Diplomarbeit, Technische Universität Ilmenau, 2009
- [132] WERNSTEDT, J.: Experimentelle Prozessanalyse., Verlag Technik Berlin, 1. Auflage, 1989
- [133] WIESENDANGER, R.: Scanning Probe Microscopy and Spectroscopy: Methods and Applications., Cambridge University Press, 1. Auflage, 1994
- [134] WILAMOWSKI, B.M. ; IPLIKCI, S. ; KAYNAK, O. ; EFE, M.O.: An algorithm for fast convergence in training neural networks., In: *International Joint Conference on Neural Networks*, Proceedings, vol. 3, pp. 1778-1782, 2001
- [135] WORDEN, K. ; WONG, C.X. ; PARLITZ, U. ; HORNSTEIN, A. ; ENGSTER, B. ; TJAHJOWIDODO, T. ; AL-BENDER, F. ; RIZOS, D.D. ; FASSOIS, D.G.: Identification of pre-sliding and sliding friction dynamics: Grey box and black-box models., In: *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 514-534, 2005
- [136] ZEHETNER, J. ; REGER, J. ; HORN, M.: A Derivative Estimation Toolbox based on Algebraic Methods - Theory and Practice., In: *IEEE International Conference on Control Applications*, Proceedings, pp. 331-336, Singapur, 2007
- [137] ZEHETNER, J. ; REGER, J. ; HORN, M.: Echtzeit-Implementierung eines algebraischen Ableitungsschätzverfahrens., In: *at-Automatisierungstechnik* 55(11), pp. 553-560, 2007
- [138] ZIEGLER, J.G. ; NICHOLS, N.B.: Optimum Settings for Automatic Controllers., In: *Transactions on the ASME*, vol. 64, pp. 759-768, 1942
- [139] ZILL, H. ; HULTZSCH, E. ; LIPPERT, E.: Messen und Lehren im Maschinenbau und in der Feingerätetechnik., VEB Verlag Technik Berlin, 3. Auflage, 1974
- [140] ZIMMERMANN, J. ; SAWODNY, O. ; HAUSOTTE, T. ; JÄGER, G.: Friction Modelling of a linear high-precision Actuator., In: *IFAC World Congress*, Proceedings, Prag, 2005
- [141] ZSCHAECK, S.: Kompensation von Störgrößen zur nanometergenauen Positionierung von kugelgeführten Parallelkinematiken., Diplomarbeit, Technische Universität Ilmenau, 2008
- [142] ZSCHAECK, S.: Untersuchung und Modellierung des Reibverhaltens von Kugelführungen im Nanometerbereich mit Hilfe eines physikalisch motivierten Reibmodells., Studienarbeit, Technische Universität Ilmenau, 2007